UNIVERSITY OF LIVERPOOL

Understanding the Nasopharyngeal Carriage Dynamics of *Streptococcus pneumoniae* and other Microbiota in Malawian Children and Adults

By

Arox Wadson Kamng'ona

(*BSc* (*Hons*), *MSc*)

Thesis submitted in accordance with the requirements of the University of Liverpool and University of Malawi for the degree of Doctor of Philosophy

January 2014

Certificate of approval

The thesis of Arox Kamng'ona is approved by the Thesis Examination Committee:

Prof. Adamson Muula

(Chairman, Post Graduate Committee)

Dr D. Everett, Prof. R. Heyderman, Prof. S. Bentley and Dr K. Gray

(Supervisors)

Prof. Neil Hall and Dr. Wilson Mandala

(Internal Examiners)

Associate Prof. Edward Senga

(Head of Department)

Author's declaration

I declare that this thesis, submitted for the degree of Doctor of Philosophy is the result of my own investigation. The work done by others has been acknowledged accordingly (see table below). The material presented here is original and has not been presented, either wholly or part thereof, for any other degree or qualification.

Institution* ***Chapter 2, Sec 2.0**	**Task**	**Responsibility**
MLW	Collection of nasopharyngeal swabs	Others
	Sample selection	Arox Kamng'ona
	Pneumococcal culturing and identification	Arox Kamng'ona
	DNA extraction	Arox Kamng'ona
	Microarray data analysis in STATA (adults)	Arox Kamng'ona
	Microarray data analysis in STATA (children)	Shared
BμG@S	Pneumococcal culturing and identification	Arox Kamng'ona
	Microarray DNA labeling and hybridisation	Arox Kamng'ona
	PCR serotyping and gel analysis	Arox Kamng'ona
	Microarray feature extraction	Others
WTSI	DNA extraction (culture independent)	Arox Kamng'ona
	16S rRNA PCR amplification	Arox Kamng'ona
	Low melt gel DNA extraction	Arox Kamng'ona
	MLST analysis	Arox Kamng'ona
	Analysis of pneumococcal *CPS* locus genes	Arox Kamng'ona
	454 sequencing	Others
	454 sequence data analysis	Shared
	Whole genome sequencing	Others
	Genetic recombination detection	Shared
Liverpool	Identifying 6B CPS locus variants	Arox Kamng'ona
	DNA extraction from 6B variants	Arox Kamng'ona
	Characterising 6B variant *CPS* genes	Arox Kamng'ona
	Testing 6B variants in mouse models	Others

SIGNED……**DATE**…….

Acknowledgements

I would like to acknowledge GOD the almighty for keeping me sane. My sincere gratitude goes to my long time friend and dear loving wife Mtisunge for her unwavering love and for sacrificing our quality time together to ensure this work was thoroughly done. And to my beautiful girls Juanita, Tamanda and Nelly, thank you for motivating me to keep working hard, I love you.

I am greatly indebted to my primary supervisor Dean Everett; and my secondary supervisors: Stephen Bentley, Robert Heyderman, and Katherine Gray. You guys worked tirelessly in your respective roles to ensure this project became a success story. I consider it an honor to have been given an opportunity to be supervised by accomplished scientists and academics like you. And to my PhD Advisory Team: Edward Senga, Neil French and Jason Hinds, thank you for making sure I was always on track and making good progress.

This work was conducted in collaboration with the Bacterial Microarray Group at St Georges (BμG@S) in London, where Jason Hinds and Katherine Gould deserve special mention for teaching me the lab work and data analysis. I also collaborated with the Sanger Institute in Cambridge for all my sequencing work and special mention go to Susannah Salter (for assisting with lab work and analysis of 16S rRNA sequence data), Allan Walker and the entire Team 81. Back to Malawi, I am greatly indebted to Naor Bar-Zeev for helping with all the statistical work. Chrispin Chaguza my friend, you made bioinformatics look easy, I will forever be thankful for your help. Chisomo Msefula, thanks for helping me understand Artemis better. Jennifer Cornick, Benard Kulohoma, Maaike Alaerts and the entire team in the JR

investigators room at MLW, thanks for always being there for me from the start of the show, and for technical discussions about my work. Kondwani Jambo, I am thankful for your help with concepts in immunology. Mavuto Mukaka, my longtime friend, thanks for helping me understand concepts of statistics necessary for my work and for helping with the formatting of this PhD thesis. I am greatly thankful for the unwavering support I received from the MLW community particularly personnel of Molecular Biology, Microbiology Laboratory and the H1N1 team for ensuring that I had everything I needed for my laboratory work. In this particular regard, Mavis Menyere, Wilson Mwafulirwa, Mike Moore and Bridget Dennis deserve special mention.

To my late dad Joseph Kamng'ona Chirwa and mum Magret Kamng'ona Chirwa (nee Banda) thank you. Your love, care and support brought me this far! Your labour was not in vain. May you rest in Gods' eternal peace dear parents.

Dedication

I dedicate this thesis to my wife Mtisunge Kamng'ona and daughters Juanita, Tamanda and Nelly.

Table of contents

List of tables

List of figures

Glossary of terms

Adr	*O*-acetyltransferase
AOM	Acute otitis media
APCs	Antigen presenting cells
ART	Anti-retroviral therapy
ATP	Adenosine tri-phosphate
BgaA	beta-galactosidase A
BHI	Brain heart infusion
BμG@S	bacterial microarray group at St George's
CbpE	Choline binding protein E
ChoP	Phosphorylcholine
COBA	Colistin oxolinic acid blood agar
CPS	Capsular polysaccharide
CR2	complement receptor 2
DHFR	dihydrofolate reductase
DHPS	dihydropteroate synthase
dNTP	Deoxyribonucleotide triphosphate
Eno	Enolase
EPI	Expanded programme on immunisation
FBS	Fetal bovine serum
GalNac	N-acetyl-D-galactosamine
GAVI	Global Alliance for Vaccines and Immunisation
HG	Homology group
HGID	homology group identification
HIV	Human immunodeficiency virus

IgA	immunoglobulin A
IGH	Institute of Global Health
IgM	immunoglobulin M
ILI	Influenza-like illness
IPD	Invasive pneumococcal disease
KPS	Karonga Prevention Study
LPXTG	Leu-Pro-any-Thr-Gly
LytA	Autolysin A
MHC II	Major histocompatibility complex class II
MLST	Multilocus sequence typing
MLW	Malawi Liverpool Wellcome Trust
MNCs	Mononuclear cells
NaCl	Sodium chloride
NanA	Neuraminidase A
NanB	Neuraminidase B
NGS	next generation sequencing
NT	Nontypeable
NVT	Non-vaccine type
OTU	Operational taxonomic unit
PAFr	platelet activating factor receptor
pathID	pathogen identification
PavA	Pneumococcal adhesion and virulence A
PBMCs	peripheral blood mononuclear cells
PBPs	penicillin binding proteins
PBS	Phosphate buffered saline

PclA	pneumococcal collagen like protein A
PCR	Polymerase chain reaction
PCV	Pneumococcal conjugate vaccine
PCV10	10-valent pneumococcal conjugate vaccine
PCV13	13-valent pneumococcal conjugate vaccine
PCV7	7-valent pneumococcal conjugate vaccine
PdgA	peptidoglycan *N*-acetylgluco- samine-deacetylase A
PiaA	Pneumococcal iron acquisition A
pIgR	polymeric immunoglobulin
PiuA	Pneumococcal iron uptake A
Ply	Pneumolysin
PPV23	23-valent pneumococcal polysaccharide vaccine
PsaA	Pneumococcal surface antigen A
PspA	Pneumococcalsurface protein A
PspC	Pneumococcal surface protein C
RFLP	restriction fragment length polymorphism
SARI	Severe Acute Respiratory Syndrome
sIgA	Secretory IgA
SP-CPS	*Streptococcus pneumoniae* capsular polysaccharide
STGG	Skim milk, tryptone, glucose and glycerine
STID	serotype identification
StrH	beta-*N*-acetylglucosaminidase
TMP-SMX	trimethoprim-sulfamethoxazole
VT	Vaccine type
WHO	World Health Organisation

WTA	Wall Teichoic Acid
WTSI	Wellcome Trust Sanger Institute

Abstract

Streptococcus pneumoniae naturally colonises the human nasopharynx, where it normally resides as a commensal. However, any change in the bacterial and host condition in the nasopharynx may lead to invasive pneumococcal disease. The human nasopharynx is also home to a broad range of other microbiota, which interact with *S. pneumonia*e, consequently impacting on its carriage dynamics. This thesis investigated the carriage of *S. pneumoniae* and other microbiota in the nasopharynx of Malawian children and adults, and factors that may impact on this. To characterise *S. pneumoniae* in carriage, a microarray was used, whereas the nasopharyngeal microbiota were characterised using 16SrRNA gene sequencing. Results demonstrated a broad range of carriage pneumococcal serotypes in Malawian children and adults. Carriage of nontypeable serotypes was higher in adults (~20%) compared to children (~1%). Based on carriage data only, PCV13 coverage was 60% in children and 30% in adults. The low PCV13 coverage demonstrates a high prevalence of non-vaccine serotypes in circulation, which may provide the scope for serotype replacement post vaccination. For the first time in Malawi, we have demonstrated that multiple carriage is not only common but also higher in children (40%) than in adults (19%). The study did not find any significant impact on serotype specific carriage or the degree of multiple carriage by HIV status. Despite the high levels of multiple carriage, there was no evidence of capsule switching in our setting. However CPS locus variants of 8 serotypes including the vaccine serotype 6B, 19A and 20 were identified. It is not known whether these variants have the improved ability to colonise or cause invasive disease, thereby necessitating further analysis. Studying other microorganisms in carriage showed that the Malawian nasopharyngeal microbiota was dominated by four phyla in both adults

and children, namely *Firmicutes* (79%), *Proteobacteria* (17%), *Actinobacteria* (3%) and *Bacteroidetes* (1%). The data demonstrated no significant difference in the distribution of phyla between adults and children, however there was broader microbial diversity amongst HIV negative subjects compared to HIV positive subjects. Specifically, HIV infection was associated with a lower prevalence of *Actinobacteria*. Although anti-retroviral therapy (ART) did not impact the nasopharyngeal microbiota, our data showed differences in phyla distribution in children's samples collected from Karonga and Blantyre. Samples from Karonga demonstrated a significantly higher carriage of *Actinobacteria* and *Proteobacteria*, while *Firmicutes* were predominant in Blantyre. This diversity could possibly be attributed to existing differences in age and HIV prevalence between the two sample groups. In conclusion, our study has demonstrated a high degree of serotype diversity and multiple carriage in the Malawian population, which suggests a greater likelihood of serotype replacement post-vaccination. This study therefore, recommends further carriage surveillance studies post PCV13 in the Malawian population. Such studies should evaluate the impact of PCV13 on carriage in the Malawian population, which in turn, will provide an estimate of herd immunity. The study has also generated important preliminary findings on dominant phyla in carriage, however further studies are recommended to evaluate the impact of other microbiota on pneumococcal carriage dynamics.

Chapter 1: Nasopharyngeal carriage of *Streptococcus pneumoniae*

1.0 Introduction

The human body is colonised by thousands of species of bacteria and a smaller number of viruses, fungi, and protozoa (Schaechter et al., 1998). These microbes are termed normal flora because they colonise the human body immediately after birth and remain throughout life (Schaechter et al., 1998). Normal flora can colonise many sites of the human body including the skin, respiratory tract (nose and oropharynx), urinary tract, digestive tract and the genital system (Figure 1.1), while areas such as the brain and the circulatory system are normally sterile (Schaechter et al., 1998). The nasopharynx is a major ecological reservoir of colonising micro-organisms which include potential pathogens such as *Streptococcus pneumoniae, Haemophilus influenzae, Moraxella catarrhalis, Staphylococcus aureus and Neisseria meningitidis* (Pettigrew et al., 2008, Nouwen et al., 2001, Bogaert et al., 2004c, Bogaert et al., 2005). Colonisation with these pathogens may occur more than once early in life in almost all humans (Faden et al., 1997, Faden et al., 1996). In industrialised countries, *S. pneumoniae, M. catarrhalis* and nontypeable *H. influenza*e can colonise as many as 54%, 72% and 33% of children, respectively, by 1 year of age (Faden et al., 1997). In developing countries such as Malawi and the Gambia, the rates of colonisation may be considerably high (Feikin et al., 2003, Kwambana et al., 2011).

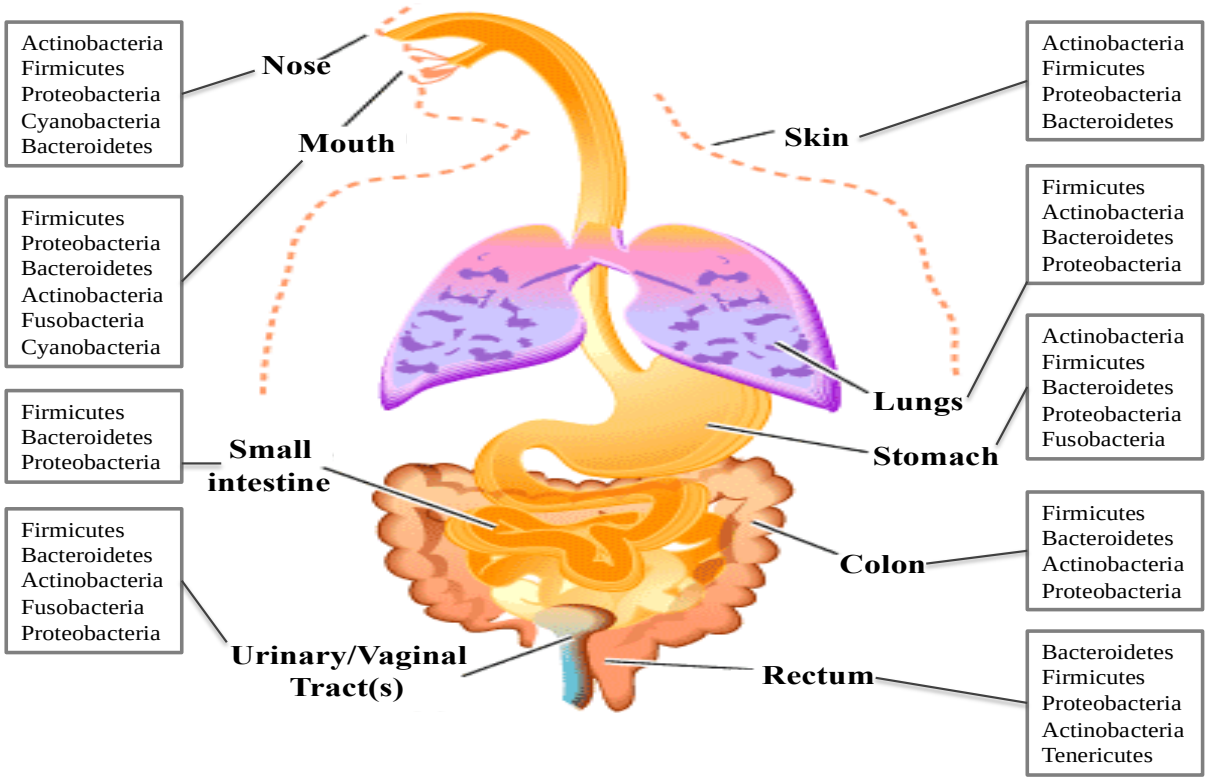


Figure 1.1 Parts of the human body colonised by normal flora

The figure lists the commonly isolated phyla from these environments (Oliver, 2003, McHardy et al., 2013, Cho and Blaser, 2012, Ahmed et al., 2007).

The colonising microbial species may be confined to their original site of attachment, where an asymptomatic carrier state is established; however, changes in the host or bacterial condition may lead to the spread of these potential pathogens to sterile sites such as the bloodstream, where they cause potentially fatal infections such as septicaemia (Kyaw et al., 2002, Nouwen et al., 2001). Colonising flora act as reservoir and source of bacterial transmission within the community and may become pathogens (Tancrede, 1992, Givon-Lavi et al., 2002). *S. pneumoniae* is an example of nasopharyngeal commensal organism, which is known to cause invasive pneumococcal disease, and is responsible for over a million deaths worldwide each year (Johnson et al., 2010, O'Brien et al., 2009), with the greatest burden experienced in sub-Saharan Africa (SSA). Thus realising the importance of carriage in the spread of pathogens and the development of disease, it is essential to undertake studies aimed at understanding carriage of such pathogens in the nasopharynx, which may play a vital role in disease control.

1.1. *Streptococcus pneumoniae* bacteriology

Streptococcal species are Gram-positive, non-motile, non-spore forming, catalase-negative organisms that are naturally found in pairs or chains (Carver et al., 2005). Their classification at the genus level is based on a number of factors such as colony morphology, haemolysis, biochemical reactions, and serologic specificity. They are divided into three groups based on the type of haemolysis on blood agar plates: beta-hemolytic, gamma-hemolytic or alpha-hemolytic (Facklam, 2002).Beta-hemolytic streptococci, which include *pyogenes* (group A) and *agalactia*e (group B), undergo complete haemolysis of red blood cells resulting in the appearance of a clear zone surrounding the colony on blood agar (Facklam, 2002). Gamma-hemolytic

streptococci do not undergo any haemolysis and include Enterococcus species such as *E. faecalis* and *E. faecium*. Alpha haemolytic streptococci are involved in partial reduction of haemoglobin (red blood cells) to methemoglobin, shown as a green or brown discoloration in the medium surrounding the colony (Facklam, 2002). *S. pneumoniae* and the viridans belong to Alpha-hemolytic streptococci (Kellogg et al., 2001). The difference between *S. pneumoniae* and viridans is that the former are mostly encapsulated and respond positively to Quellung reaction while all of the latter do not have a polysaccharide capsule and they yield a negative result to the Quellung test (www.cdc.gov). Another distinguishing feature is that pneumococcal strains are bile soluble and sensitive to optochin, while viridans are bile insoluble and optochin resistant (Wasilauskas and Hampton, 1984). Pneumococcal cells are surrounded by a cell wall, which protects the bacterial cell from bursting by the internal pressure, thereby maintaining the structural integrity of the cell (Bui et al., 2009). The main constituents of pneumococcal cell wall are peptidoglycan and the covalently attached wall teichoic acid (WTA) (Vollmer et al., 2008). The pneumococcal cell wall is surrounded by a polysaccharide capsule which renders a mucoid (smooth) appearance to colonies on agar (Todar, 2010). The polysaccharide capsule also defines pneumococcal serotypes based on their reaction to factor typing sera. These sera are specific to immunochemical differences that exist between the pneumococcal capsular polysaccharides. Based on differences in the immunological properties of the polysaccharide capsule, 95 serotypes have been reported (Park et al., 2007, Bratcher et al., 2010, Oliver et al., 2013). Serotypes are clustered into 46 sero-groups based on the heterogeneous structural and chemical composition, and on antigenic properties of capsular polysaccharides resulting from biosynthetic locus arrangements (Bentley et al., 2006). Pneumococcal strains that lack the

polysaccharide capsule have also been identified. These pneumococcal strains have usually been described as strains which are optochin sensitive and bile soluble but that cannot be typed with either antisera or the multiplex PCR (Shayegani et al., 1982; Hathaway et al., 2004; Hanage et al., 2006), hence popularly known as nontypeable (NT) pneumococcal strains. Analysis with microarray serotyping (Chapter 2, section 2.6) revealed that NT strains may have a genotype that matches the CPSgene content of known serotypes, resembles remnants of serotypes, contains no known CPSgenes or identifies them as non-*pneumoniae Streptococcus* species (Hinds et al, 2009). Recently, whole genome sequencing and analysis of the genes between *dexB* and *aliA* genes has classified NTs into three groups based on whether there is complete deletion of the CPS locus (group NT1), presence of putative novel surface protein *nspA* gene at the CPS locus instead of CPS genes (group NT2) or those containing a well-conserved *aliB* gene cluster (group NT3) (Salter et al., 2012).

1.2. Nasopharyngeal carriage of *S. pneumoniae*

By definition, the time when a pneumococcal strain establishes itself within a host is known as acquisition, whereas persistent colonisation of the nasopharynx is referred to as carriage(Simell et al., 2012a). Nasopharyngeal carriage is a major factor in the pathogenesis of pneumococcal disease and is responsible for the spread of the pathogen within a given population. It is reported to occur early in life and persists in all age groups. Pneumococcal carriage is also reported to be more prevalent in children than adults, with colonisation rates increasing from birth to reach a peak (>80% in some populations) at the age of 1-2 years, and decreasing with age thereafter to less than 10% (Lloyd-Evans et al., 1996, Hussain et al., 2005, Abdullahi et al., 2012a, Hill et al., 2006, Regev-Yochay et al., 2010, Abdullahi et al., 2008a).

Pneumococcal carriage rates are reported to be higher in the developing regions with countries such as Malawi (Feikin et al., 2003), Gambia (Lloyd-Evans et al., 1996, Hill et al., 2006) and Nigeria (Adetifa et al., 2012) reporting carriage rates of more than 80%in children. In sub-Saharan Africa, the average carriage rate of 63.2% was shown in children less than 5 years old, decreasing with age to 28% in those older than 15 years (Usuf et al., 2014). In the developed countries, carriage rates are relatively lower compared to the developing countries. In Sweden, carriage rates of up to 32% were reported in children less than 2 years of age (Aniansson et al., 1992); while in the UK a 52% carriage rate was reported in children of the same age range (Hussain et al., 2005).

The duration of carriage for the pneumococcus varies and periods of up to 40 weeks have been reported (Sleeman et al., 2006, Turner et al., 2012, Gratten et al., 1986). The persistence of a serotype in carriage is associated with the biochemistry of a polysaccharide capsule, where serotypes with thicker capsules are reported to persist for longer durations in carriage compared to serotypes with thinner capsules (Hogberg et al., 2007, Weinberger et al., 2009). Such heavily encapsulated serotypes have also been shown to exhibit greater ability to escape neutrophil-mediated killing, which may prolong their duration in carriage (Weinberger et al., 2009). Other host factors such as serotype-specific and CD4+ Th17 cells mediated non-serotype specific acquired immune responses have also been reported to play a role in the dynamics of pneumococcal carriage in the nasopharynx (Cobey and Lipsitch, 2012, Malley et al., 2005). Other reports indicate that less immunogenic serotypes tend to be carried for a much longer period of time than the more immunogenic serotypes (Rosen et al., 1984, Obaro and Adegbola, 2002). Carriage is also reported to be

dependent on seasons with peak rates occurring in the rainy and cold seasons (Gray et al., 1980, Abdullahi et al., 2012a). Other risk factors for pneumococcal carriage include childcare attendance, number of siblings in the home and health status and co-colonisation with other pathogens (Huang et al., 2009, Syrjanen et al., 2001, Regev-Yochay et al., 2004a, Bogaert et al., 2001, Abdullahi et al., 2008b).Exposure to viral infections in the upper respiratory tract has been reported to be a predisposing factor to bacterial colonisation. Viruses are thought to disrupt the respiratory mucosal epithelium thereby enhancing bacterial adhesion to respiratory epithelial cells (Glennie et al., 2010). In addition, inflammatory response to viral infection is thought to up-regulate the expression of molecules utilised by bacteria as receptors, eventually promoting adhesion (Jacoby et al., 2007, Peltola and McCullers, 2004). A recent study has shown a significant association between viral infection and detection of *S. pneumoniae* in carriage in children (Rodrigues et al., 2013a). The same study also reported an interesting finding where detection of *H. influenzae* and *H. influenzae* colonisation density was positively associated with rhinitis symptoms in an age independent manner (Rodrigues et al., 2013a). Since rhinitis is a viral infection, the finding has led to the hypothesis that *H. influenzae* promotes its own transmission within a given population by inducing or amplifying rhinitis in children, which could suggest that carriage may not necessarily be asymptomatic (Rodrigues et al., 2013a).

1.2.1. Multiple carriage of *S. pneumoniae*

Multiple carriage is the simultaneous carriage of at-least two serotypes in the nasopharynx, and is thought to promote horizontal gene transfer (Hiller et al., 2010, Thomas and Nielsen, 2005). By promoting horizontal gene transfer, multiple carriage

has potentially important clinical implications in a given population as it may facilitate the rapid development of antibiotic resistance and vaccine escape through capsule switching (Marks et al., 2012, Hiller et al., 2010, Ding et al., 2009, Donkor et al., 2011). Capsule switching involves a change of serotype of a single clone by alteration or exchange of its CPS locus (Wyres et al., 2013). Multiple carriage also impacts the dynamics of pneumococcal carriage through bacterial interference (Shak et al., 2012). Multiple serotype carriage (co-colonisation) occurs when more than one distinct pneumococcal strains occupy a particular niche at the same time (Brugger et al., 2009).

Data on the actual rates of multiple carriage of pneumococcal serotypes in the nasopharynx are very limited, however, previous estimates ranged from 1.3 % to 20% in HIV uninfected adults and children (Gratten et al., 1994, Huebner et al., 2000a, Hare et al., 2008, Rivera-Olivero et al., 2009, Brugger et al., 2009, Kaltoft et al., 2008). With improved serotyping techniques, however, much higher carriage rates (48%, n=129) have recently been reported (Turner et al., 2011). Multiple carriage in the nasopharygnx could facilitates horizontal gene transfer between pneumococcal strains (Marks et al., 2012). By promoting horizontal gene transfer, multiple carriage has potentially important clinical implications in a given population as it may facilitate the rapid development of antibiotic resistance (Croucher et al., 2011)and vaccine escape through capsule switching (Brugger et al., 2009). Capsule switching involves a change of serotype of a single clone by alteration or exchange of its CPS locus (Wyres et al., 2013).

Pneumococcal strains are reported to be highly transformable; as such they are 50 times more likely to undergo horizontal genetic transfer than genetic mutations (Feil et al., 2000, Claverys et al., 2009). The occurrence of capsule switch has been observed in various settings (Scott et al., 2011, Croucher et al., 2013) including Malawi (Everett et al., 2012), and this has implications for the efficacy of capsular based polysaccharide vaccines (Brueggemann et al., 2007b). The data on multiple serotype carriage prevalence in a given population could provide a basis for understanding and even predicting the emergence of new strains under vaccine selection pressure (Brugger et al., 2009, Mulholland and Satzke, 2012), especially for conditions involving multiple carriage of vaccine and non-vaccine serotypes. Multiple carriage data could also be used to explain whether non-vaccine serotypes emerging post vaccination are a result of true serotype replacement or unmasking of other serotypes (Figure 1.2). Under true serotype replacement, the unvaccinated populationmay initially carryboth vaccine and non-vaccine types, while vaccine recipients may carry a reduced number of vaccine serotypes and non-vaccine serotypes not previously detected. Under unmasking, by contrast, the unvaccinated carry non-vaccine serotypes below detectable levels and vaccine serotypes. Post-vaccination, the low abundant non-vaccine types may become dominant serotypes (Figure 1.2) (Lipsitch, 1999). Unmasking is often difficult to detect if only a single colony is sampled for serotype analysis (Lipsitch, 1999).

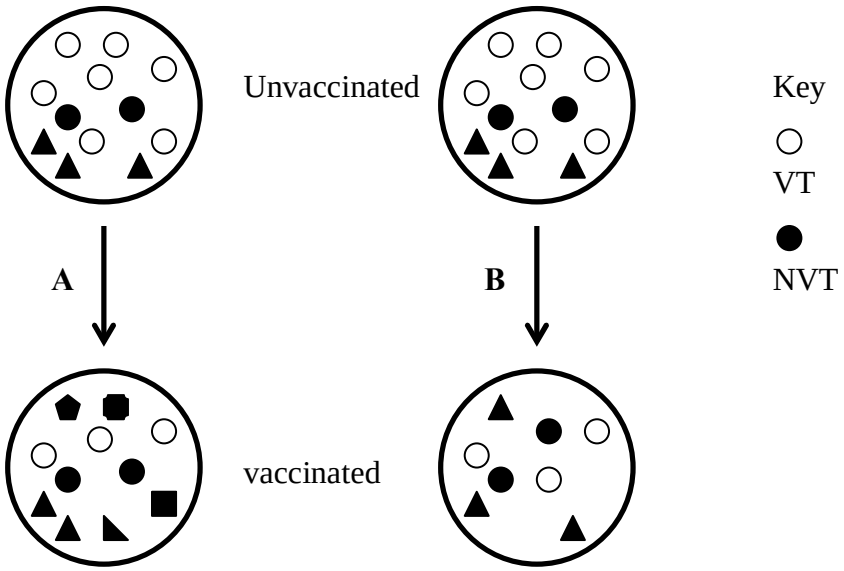


Figure 1.2 Serotype replacement model

A model depicting true serotype replacement (A) and unmasking (B) in carriage under vaccine selection pressure. Large circles represent plated samples from controls (unvaccinated) and vaccine recipients (vaccinated). The left side shows true serotype replacement, while the right side shows unmasking. Serotypes covered by vaccine (VT) are shown by open circles, while serotypes not covered by the vaccine (NVT) are colour coded black (Lipsitch, 1999).

1.2.2. Detection of multiple carriage

A major barrier to an accurate estimation of the prevalence of multiple pneumococcal carriage is lack of standard methods of detection. Conventionally, culture based techniques and serotyping of individual colonies have been employed (Kaltoft et al., 2008). Although attempts to increase the sensitivity of culture based methods by combining with molecular techniques such as multiplex PCR have been made (Rivera-Olivero et al., 2009), this approach still has limited practical application for very large numbers of samples (Huebner et al., 2000a) and it is biased towards viable and most abundant serotypes, as they are dependent on picking a

single colony from an agar plate either for sub-culturing or PCR. Ideally, culture independent methods of detecting co-colonisation would give a more accurate picture of multiple carriage. A culture independent technique for detection of co-colonisation based on PCR amplification of a non-coding region adjacent to the pneumolysin gene (*plyNCR*) has been reported (Brugger et al., 2009). The authors stated that no amplification was observed when the PCR was conducted on other streptococcal and non-streptococcal species, which suggested pneumococcal specificity, however the data was not shown. Using restriction fragment length polymorphism (RFLP) and terminal RFLP analysis of the *plyNCR* region, the numbers and relative abundance in co-colonising strains were determined (Brugger et al., 2009). Although the sensitivity of the technique was slightly higher (50% vs 40%) than culture based methods, it only works best for samples with at most two strains and differences in RFLP detected by one restriction enzyme only (Brugger et al., 2009). Thus there is still need for novel methods to fully understand the prevalence of multiple carriage in various settings.

1.2.3. Pneumococcal carriage and disease

The evidence of carriage as a predisposing factor and a first step in pneumococcal disease has been demonstrated in animal models, where nasal inoculation with pneumococcal strains resulted in development of acute otitis media (AOM) (Peltola et al., 2006, van der Ven et al., 1999) and/or invasive disease (Kadioglu et al., 2000, Ogunniyi et al., 2007b). In humans, very few studies have been conducted to establish this association, with the predominant infection being AOM. Development of AOM was associated with serotypes newly acquired (Gray et al., 1980), and not necessarily serotypes established in carriage. A recent study demonstrated that

human carriage of *S. pneumoniae* was associated with protection against re-carriage and invasive pneumococcal disease. This was attributed to an increase in both mucosal and serum IgG levels to pneumococcal proteins and polysaccharide, which resulted in an increase in opsonophagocytosis (Ferreira et al., 2013). It has been reported that most serotypes cause AOM at a frequency that is proportional to their prevalence in carriage (Hanage et al., 2004), suggesting a strong association between carriage and the onset of AOM. The most common serotypes causing AOM worldwide are 3, 6A, 6B, 9V, 14, 19A, 19F and 23F (Rodgers et al., 2009). The association between carriage and invasive disease, however, is complex, i.e., there is a discordant relationship in the distribution of serotypes between carriage and invasive disease. For example, serotypes 1, 5 and 7F are frequently isolated from invasive disease particularly in Africa, Asia and Latin America (Hausdorff et al., 2000) and often cause disease outbreaks, and yet they are rarely detected in carriage (Hanage et al., 2005, Brueggemann et al., 2003, Almeida et al., 2013). The ability to cause disease in this case is not determined by the frequency of serotypes in carriage but rather by the invasive capacity of a given serotype, which is generally stable across the globe (Greenberg et al., 2011). Further evidence for the association between acquisition and the onset of disease can be seen in children. It has been reported in children, that IPD peaks from 6-11 months, which coincidentally is the time when high rates of new pneumococcal acquisitions in carriage are experienced (Simell et al., 2012a). Finally, carriage is much higher in children in developing than developed regions, with IPD following a similar trend, further highlighting the role of pneumococcal serotype acquisition in disease. In adults, the association between carriage and invasive disease is not very clear, i.e. there is high IPD incidence at a time when the prevalence of pneumococcal carriage is very low, particularly in the

elderly (Flasche et al., 2011a, Hussain et al., 2005). Thus the susceptibility of adults to IPD may largely be influenced by other medical or social-demographic factors and not necessarily carriage prevalence. This suggests that the development of invasive pneumococcal disease cannot reliably be predicted by the prevalence of carriage. Further information is therefore still required to establish the relationship between carriage and disease. The key is with the development of novel serotyping techniques such as microarray (Turner et al., 2011), which have the ability to detect multiple carriage as well as accounting for less abundant serotypes.

1.3. Infections caused by *S. pneumoniae*

The pneumococcus is reported to be the leading cause of bacterial pneumonia, meningitis, and sepsis, which contribute substantially to high morbidity and mortality worldwide (O'Brien et al., 2009). The highest risk groups for pneumococcal disease are children less than 5 years old, the elderly, and persons with chronic underlying conditions, such as HIV/AIDS (WHO., 2007). It is estimated that pneumococcal infections are responsible for ~1.6 million deaths annually. In children under the age of 5 years alone, deaths caused by *S. pneumoniae* range from ~7000000 to ~1 million every year world-wide (2007). The highest pneumococcal mortality rates are reported in Africa, with sub-Saharan Africa (SSA) bearing the greatest burden (O'Brien et al., 2009) (Figure 1.3).

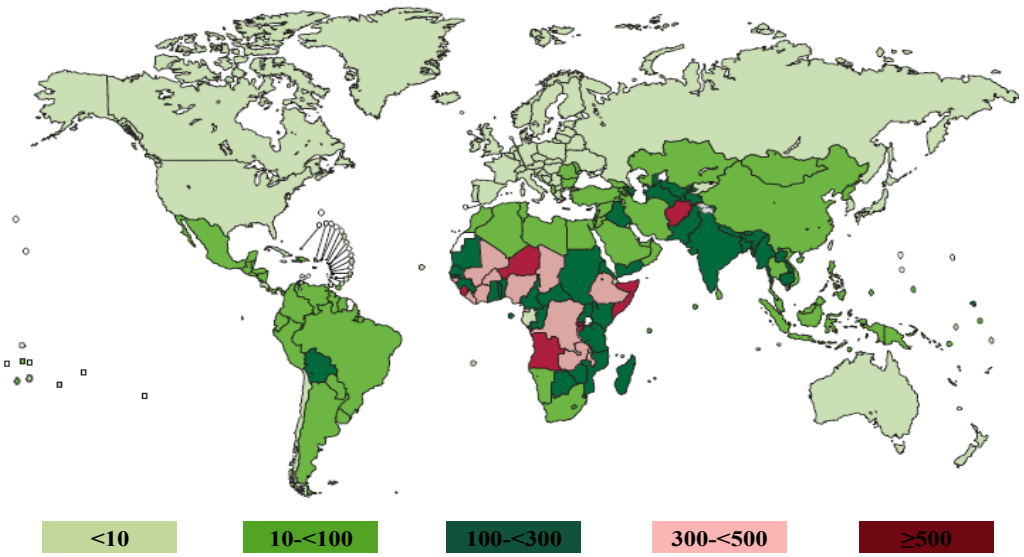


Figure 1.3 Global pneumococcal mortality rate

Pneumococcal deaths in children aged 1–59 months per 100 000 children younger than 5 years (HIV-negative pneumococcal deaths only) (O'Brien et al., 2009).

1.3.1. Global invasive serotype distribution

The global distribution of serotypes causing invasive disease shows that serotype 14 is the most commonly isolated serotype in children under the age of 5 years (Johnson et al., 2010). The five most common global serotypes accounting for approximately 60% of all serotypes causing invasive disease in children <5 years are 14, 6B, 1, 19F and 23F (Figure 1.4). Serotype 1 was more prevalent among children 2-<5 years of age than among the <2 years age group (Johnson et al., 2010, Pearce et al., 2002)

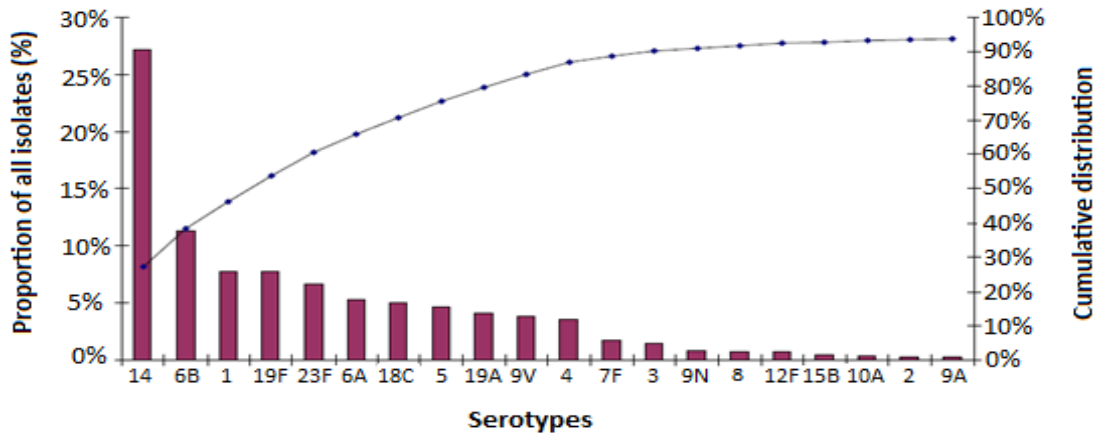


Figure 1.4 Global distribution of invasive serotypes in children

The bar graph represents the percentage of each serotype out of the total number of isolated strains in children aged less than 5 years (Johnson et al., 2010). The line graph represents the cumulative frequency, which shows that serotypes 14 through 9A account for approximately 95% of all serotypes causing invasive disease in children under the age of 5 years globally.

Published data for invasive serotype and serogroup distributions in Europe, USA and Asia are similar, although minor differences in the prevalence of other serotypes exist. In the Netherlands, the most prevalent serotypes in children are 19F, 6B, 6A, 9V, and 23F (Bogaert et al., 2001). In the USA and Vietnam, the most commonly isolated serotypes were 6B, 14, 19F, 23F (Yeh et al., 2003) and 19, 23, 14, 6, and 18 (Parry et al., 2000) respectively. There is a lot of overlap between patterns of serotype distribution in Africa and other settings (Table 1.1). There has also been no major differences in serotype distribution between children with risk factors such as day care centre attendance (Bogaert et al., 2001) or upper respiratory tract infections and healthy children (Syrjanen et al., 2001, Marchisio et al., 2003), although respiratory tract infections are known to impact on carriage (Valles et al., 2006).

Table 1.1 Common carriage and invasive serotypes in African children

Reference publication	Year	Country	Age-Group	Isolate	Serotypes/serogroups
(Woolfson et al., 1997)	1994	Zambia	≤6 years	carried	14*, 19, 23
(Huebner et al., 2000b)	2000	RSA	≤5 years	carried	6B, 19F*, 6A*, 23F, 14, 19A
(Adegbola et al., 2006)	1996-2003	Gambia	≤6 years	invasive	1, 5,6A, 14, 19A, 23F
(Hill et al., 2008a)	2008	Gambia	≤1 year	carried	6B, 19F, 6A, 14, 23F*
(Joloba et al., 2001)	1995	Uganda	≤3 years	carried	19, 6, 23, 9, 14
(Morais et al., 2007)	2003-2006	Mozambique	≤15 years	invasive	1, 5, 6A, 6B, 19A, 19F, 14, 23F
(Valles et al., 2006)	2003	Mozambique	≤5 years	carried	19F, 19A, 23F, 6A, 6B*
(Huebner et al., 1998)	1998	Botswana	≤5 years	carried	14, 19F, 19A, 6A, 6B, 4

**The five most common carriage serotypes in sub-Saharan Africa (Usuf et al., 2014), which also rank among the top seven serotypes causing invasive disease in children globally (Johnson et al., 2010).*

1.4. Impact of HIV infection on carriage and invasive disease

1.4.1. Global estimates of HIV infection

Globally the total number of people living with HIV is estimated to be 35.3 million, with 2.3 million new cases reported in 2012 (Figure 1.5). Sub-Saharan Africa bears the greatest burden, with ~70% of the global population of HIV infected people living in this region, and reported 1.6 million new HIV infection cases in 2012 against 2.3 million world-wide (UNAIDS, 2013a, UNAIDS, 2013b). South Africa has the largest number (4.3 million) of people living with HIV in the world, while at 24.8%; Swaziland has the highest rate of HIV prevalence in the world. In Malawi, about 1.7 million people live with HIV infection representing 10.6 % of the entire Malawian population, which is estimated to be 16 million people. The mortality rate due to HIV infection in Malawi is still high although there was a drop from 76, 000 deaths in 2005 to 43, 000 in 2011. This decrease is attributed to the scale-up of the ART program in Malawi (UNAIDS, 2012b).

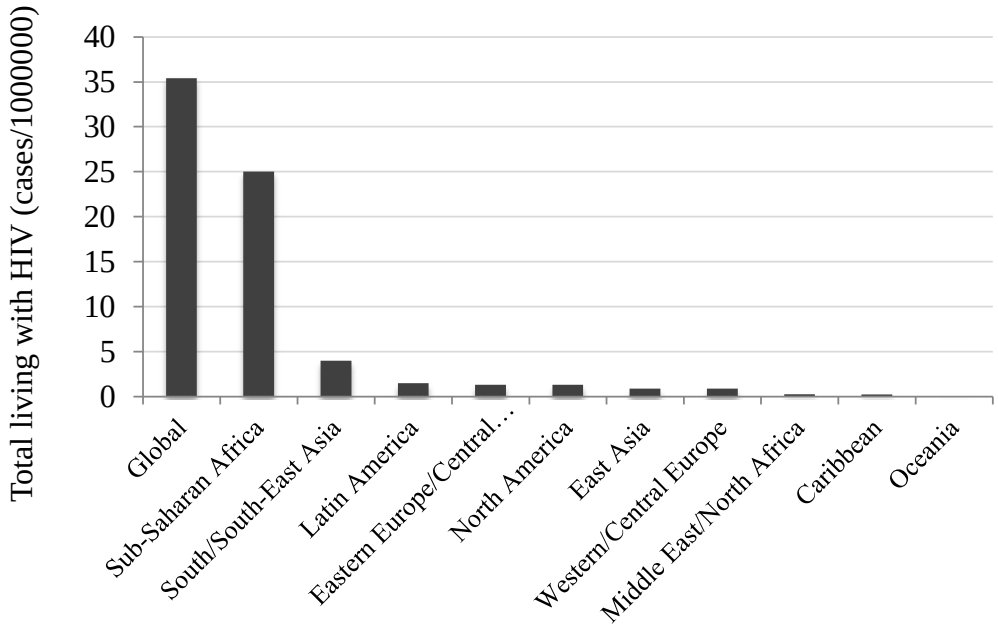


Figure 1.5 HIV prevalence by region

The region with the highest burden of HIV infection is sub-Saharan Africa, which includes Malawi (UNAIDS, 2012b, UNAIDS, 2013a, UNAIDS, 2013b).

1.4.2. Pneumococcal carriage and disease in the context of HIV infection

Risk factors for community acquired pneumococcal infections and asymptomatic carriage are numerous and diverse. These include extremes of age, low socioeconomic status, malnutrition, and conditions which compromise the host immune system such as HIV and AIDS (Farr et al., 2000b, Farr et al., 2000a, Almirall et al., 2008). HIV infection has emerged as a major global risk factor for invasive pneumococcal disease, particularly in adults (Domingo et al., 2009, Gilks et al., 1996, Jordano et al., 2004). It has significantly contributed to an increasing burden of pneumococcal disease in the developing world, particularly SSA (Zar, 2004, Nunes et al., 2011a, Gordon et al., 2002). In the USA, persons with HIV accounted for 54.2% of patients (n=399) aged between 18 to 64 years who had pneumococcal disease (Nuorti et al., 2000). In Kenya, HIV seropositivity conferred a

relative risk of 17.8 for pneumococcal infection, among female sex-workers (Gilks et al., 1996). In South Africa, similar patterns in children and adults have been reported, where HIV infection was a predisposing factor for higher rates of pneumococcal disease (Madhi et al., 2007a, Karstaedt et al., 2000, Karstaedt et al., 2001). It has also been shown that HIV-infected children and adults are susceptible to getting infected with pneumococcal strains that are resistant to penicillin treatment (Karstaedt et al., 2001, Madhi et al., 2007a), thus frustrating efforts to control pneumococcal infections in this group. Similarly, in Malawi, a high HIV prevalence has been reported among children and adult patients with invasive pneumococcal disease (Gordon et al., 2002, Carrol et al., 2007).

Since pneumococcal carriage precedes disease, several studies have been done to determine the impact of HIV infection on carriage. A higher pneumococcal carriage rate has been described in a longitudinal cohort study of Zambian mothers, where HIV infection increased the risk of colonisation (Gill et al., 2008b). In South Africa, HIV infected children had a higher prevalence of colonisation by *S. pneumoniae* and *H. influenzae* (71.6% and 74.1%, respectively) than did HIV uninfected children (50.9 and 52.0% respectively) (Madhi et al., 2007a). Results from a study in USA showed that HIV-infected subjects were more likely to be persistent pneumococcal carriers and to carry antibiotic resistant strains than were non-HIV infected subjects (Rodriguez-Barradas et al., 1997). However, findings from other studies in the USA showed that the rate of pneumococcal colonisation in HIV–infected children was equal to that of HIV-negative controls (20% vs 19%) (Polack et al., 2000). Similarly, in Brazil, there was no significant correlation between carriage in adults and HIV status, irrespective of viral load and CD4 lymphocyte count (Nicoletti et al., 2007).

The fact that different studies report different findings suggests that the relationship between HIV infection and pneumococcal carriage is very complex and therefore poorly understood. These discrepancies could be attributed to varying levels to which the functionality of the immune system is compromised by HIV infection. This was demonstrated in a recent study in Malawian adults where carriage of pneumococcal strains was marginally increased in asymptomatic HIV positive adults, whereas, a significant increase in carriage rates was observed in symptomatic HIV infection (HIV negative-13%, asymptomatic HIV positive-19% and symptomatic HIV positive-38%) (Glennie et al., 2012a). Such elevated carriage rates in adults were maintained following immune reconstitution by antiretroviral therapy (Glennie et al., 2012a). It is clear that the available data on the impact of HIV on pneumococcal carriage are inconclusive. These studies, however, did not take into account the presence of other microbiota, which are in constant interaction with *S. pneumoniae* and may have an impact on pneumococcal carriage dynamics (Bosch et al., 2013). This therefore underscores the need for further investigations.

1.5. Pathogenesis of pneumococcal infections

The pneumococcus is acquired through contact with infected droplets and its natural habitat is the nasopharynx of the human host (Figure 1.6). The pneumococcus can cause otitis media, sinusitis, and pneumonia by spreading through the larynx to the lower respiratory tract, and then in some cases in blood to cause septicaemia (Shak et al., 2012). From blood, the pneumococcus may pass through the blood brain barrier to meninges causing meningitis (Obaro and Adegbola, 2002).

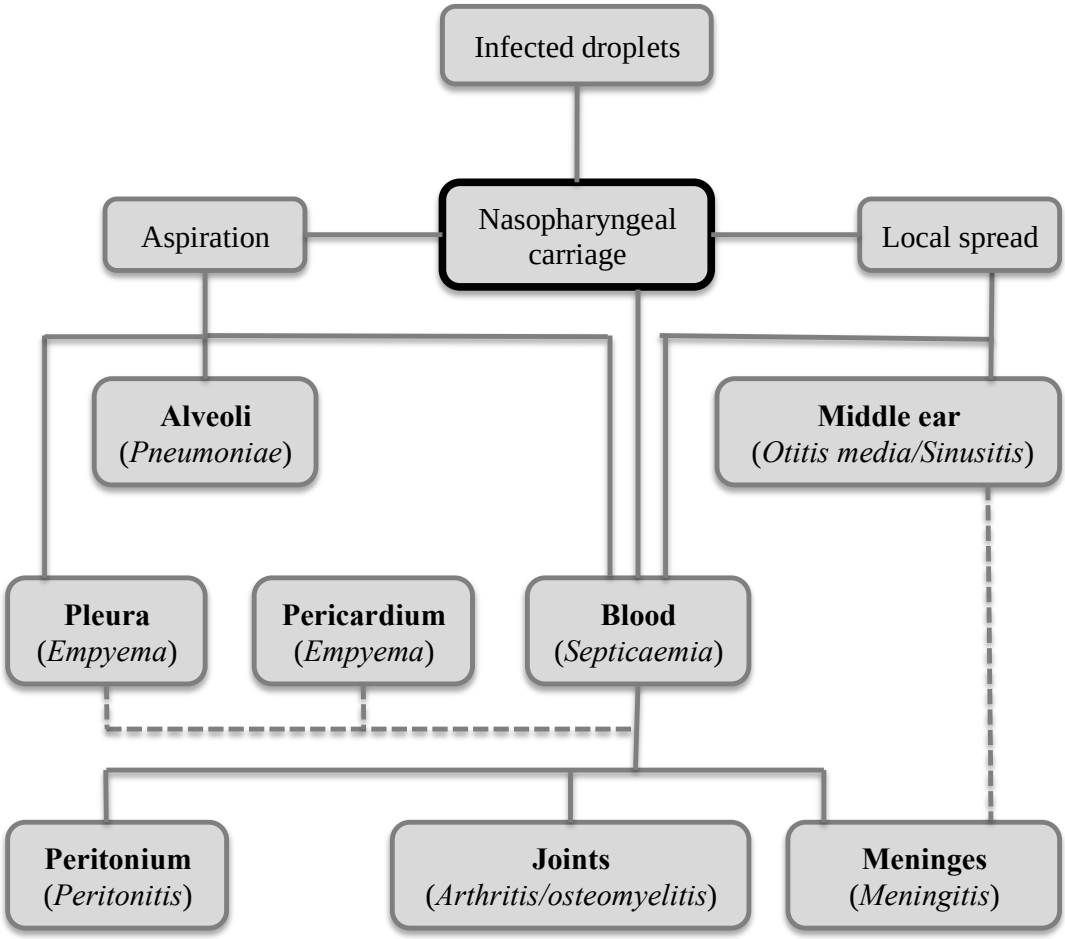


Figure 1.6 Pathogenic route for pneumococcal infection

The pneumococcus is acquired through contact with infected droplets. Upon successful colonisation in the nasopharynx, the pneumococcus can be cleared, spread locally to cause mild infections such as sinusitis or otitis media, or it can invade lung, brain and blood to cause life-threatening diseases such as pneumonia, meningitis and septicaemia respectively (Bogaert et al., 2004a).

To cause infection, the pneumococcus is initially dependent on the attachment of pneumococcal surface receptors to the host mucosal lining, invasion through internalisation of the pneumococcal strains into endothelial cells and evasion of the host immune system (Gillespie and Balakrishnan, 2000). This is achieved by expressing a wide range of virulence factors (Figure 1.7). The widely studied virulence factors include the polysaccharide capsule and both surface proteins and secreted proteins (Jedrzejas, 2001, Kadioglu et al., 2008). The pneumococcal polysaccharide capsule is the outer-most component of encapsulated pneumococcal strains, covalently attached to the cell wall peptidoglycans (Figure 1.7). The capsule of serotype 3 however, binds to peptidoglycan through a different mechanism other than covalent attachment (Sorensen et al., 1990). Unlike other capsule types, the synthesis of serotype 3 capsule does not involve formation of repeat units (Cartee et al., 2000) and occurs via a synthase dependent pathway (Cartee et al., 2000). Pneumococcal proteins such as surface exposed choline binding proteins, ATP-binding cassette transporter, metal binding proteins, LPXTG anchored proteins, pneumolysin, autolysin, and enolase play various roles in pneumococcal pathogenesis as described (Table 1.2), but the function of most pneumococcal proteins is still unknown.

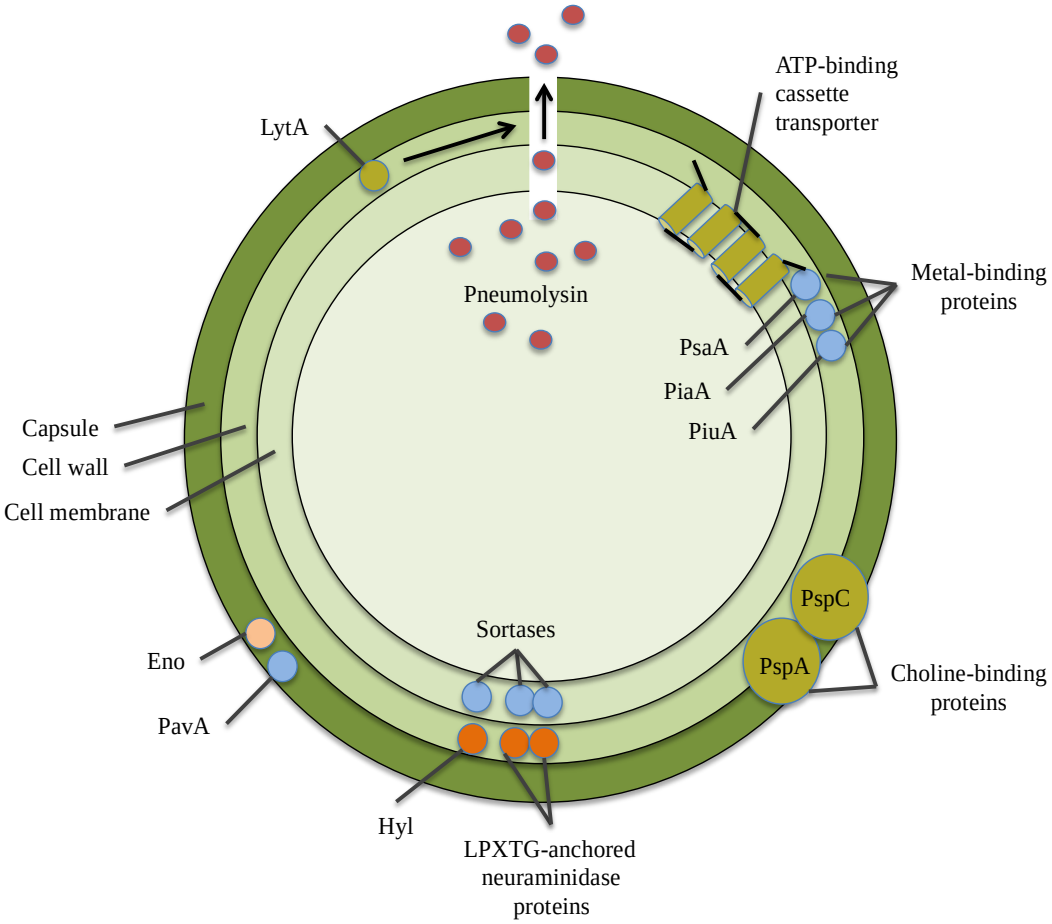


Figure 1.7 The cell of *S. pneumoniae* showing virulence factors

The capsule; the cell wall; choline-binding proteins; pneumococcal surface proteins A and C (PspA and PspC); the LPXTG-anchored neuraminidase proteins; hyaluronate lyase (Hyl); pneumococcal adhesion and virulence A (PavA); enolase (Eno); pneumolysin; autolysin A (LytA); and the metal-binding proteins pneumococcal surface antigen A (PsaA), pneumococcal iron acquisition A (PiaA) and pneumococcal iron uptake A (PiuA) (Kadioglu et al., 2008).

Table 1.2 Pneumococcal virulence factors and function

Pneumococcal Virulence Factor	Main role in colonisation and disease
Polysaccharide capsule (Nelson et al., 2007b, Hammerschmidt et al., 2005)	Prevents entrapment in the nasal mucus, thereby allowing access to epithelial surfaces. Also inhibits effective opsonophagocytosis.
Metal-binding proteins	
PsaA (Kadioglu et al., 2008)	Component of the ABC transport system, which is involved in resistance to oxidative stress
PiaA and PiuA (Kadioglu et al., 2008)	Component of the ABC transport system acquiring iron for bacterial growth and virulence
Choline-binding proteins	
PspC (Iannelli et al., 2004)	Binds to human secretory component on a polymeric Ig receptor during the first stage of translocation across the epithelium.
PspA (Rosenow et al., 1997)	Prevents binding of C3 onto pneumococcal surface. Also binds lactoferrin
Chop (Weidenmaier and Peschel, 2008)	Binds to PAFr on the epithelial surface of the human nasopharynx
LPXTG-anchored proteins	
NanA and NanB (Jedrzejas, 2001)	Contributes to adherence, removes sialic acids on host glycopeptides and mucin to expose binding sites
Other proteins	
Pneumolysin (Hirst et al., 2004, Zhang et al., 2007)	Cytolytic toxin that also activates complement, cytokine and chemokine production.
Enolase (Kadioglu et al., 2008)	Binds to fibronectin in host tissues
PavA (Holmes et al., 2001)	Binds to plasminogen within host tissues
Autolysin (Kadioglu et al., 2008)	Digests the cell wall, resulting in the release of pneumolysin

1.5.1. Pneumococcal adherence to the host mucosal surface

Respiratory mucus and lysozyme act as the first line of defence against attachment of the pneumococcus to host epithelial cells (Coonrod et al., 1991). The pneumococcus employs several different mechanisms to overcome mucus entrapment and clearing from the mucosal surface. The negative charge present on the pneumococcal polysaccharide capsule is thought to repulse the negatively charged sialic acid residues of the mucus; subsequently minimizing the chance of entrapment (Nelson et al., 2007a). The pneumococcus alters the viscosity of the mucus by de-glycosylating mucus glycoconjugates. De-glycosylation of the membrane mucus is facilitated by an enzymatic process involving neuraminidase A (NanA), neuraminidase B (NanB), beta- galactosidase A (BgaA), and beta-*N*-acetylglucosaminidase (Vestrheim et al.) (King et al., 2006a). The epithelial cilia beating may prevent the pneumococcus from binding by clearing it together with the mucus. The pneumococcus utilises pneumolysin (Ply), a pore-forming toxin to impair and decrease the beating of the ciliated cells on the mucosal epithelium, consequently allowing the pneumococcus to effectively bind to epithelial cells (Figure 1.8A) (Feldman et al., 2002, Feldman et al., 1990).

Lysozyme is a muramidase enzyme, which targets and cleaves acetylated peptidoglycan molecules (Davis et al., 2008). The pneumococcus also has to overcome the action of lysozyme released by the host to effectively attach to epithelial cells. Peptidoglycan is a polymeric structure of sugars and amino acids present in the cell wall of the pneumococcus and several other microorganisms (Davis et al., 2008). The pneumococcus expresses two enzymes, peptidoglycan *N*-glucosamine-deacetylase A (PdgA) and an *O*-acetyltransferase (Adr), which

deacetylates peptidoglycan molecules on the surface of the pneumococcus, conferring resistance to lysozyme destruction on the bacterial cell wall (Figure 1.8B) (Davis et al., 2008).

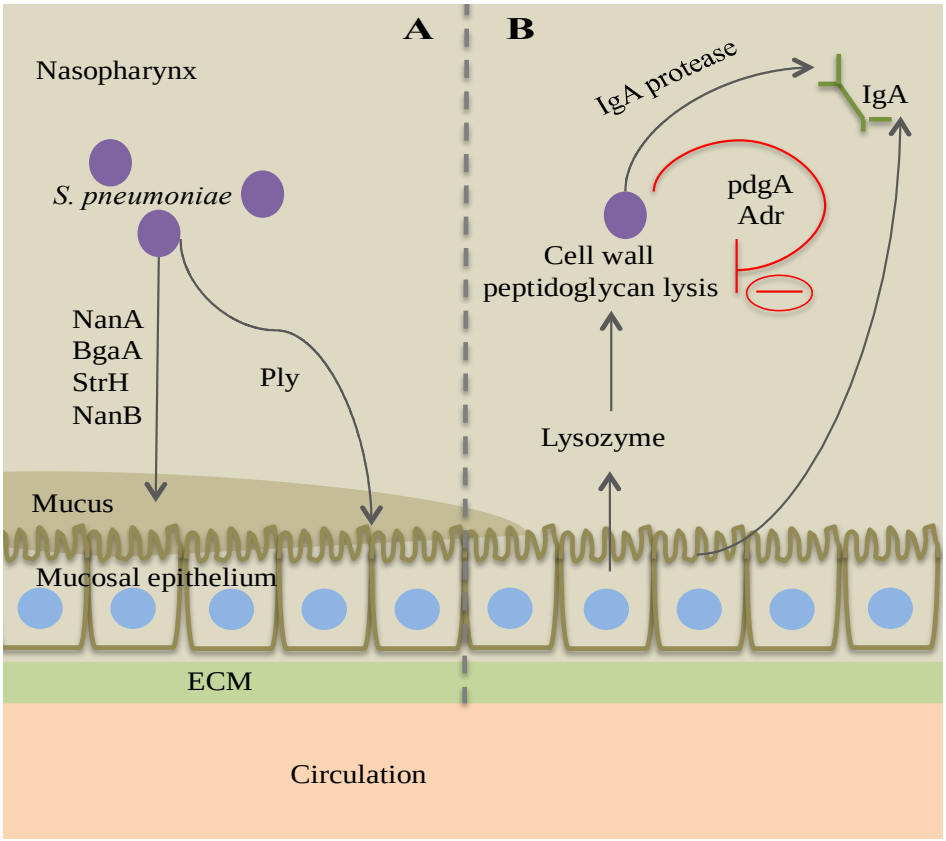


Figure 1.8 Pneumococcal attachment to host mucosal epithelium

(A) Initiation of attachment: breakdown of the mucus lining the epithelium using NanA, BgaA, StrH and NanB enzymes. This is complemented by pneumolysin (Ply), which interferes with the cell ciliary beating, thereby promoting adherence to epithelium. (B) The host releases lysozyme, a proteolytic enzyme that degrades peptidoglycans of the pneumococcal cell wall and secretory IgA. The pneumococcus employs pdgA and Adr enzymes to inhibit lysozyme, and IgA protease to destroy secretory IgA (Mook-Kanamori et al., 2011).

1.5.2. The role of host immunity in pneumococcal pathogenesis

Defence against pathogens in the respiratory tract relies upon the immune mechanism located in the airways and alveolar spaces (Asahi et al., 2002, Phalipon and Corthesy, 2003), which involves both innate and adaptive immunity (Jambo et al., 2010). The innate immune response to colonisation by *S. pneumoniae* employs factors such as C-reactive protein (CRP), an acute-phase protein that is elevated during inflammation (Du Clos and Mold, 2004, Gould and Weiser, 2002). CRP activates the complement system and also plays an important role in opsonisation and inhibition of bacterial attachment to epithelial cells (Gould and Weiser, 2002), thereby regulating pneumococcal colonisation to mucosal surfaces. The innate immune system also employs an acute inflammatory response by releasing neutrophils and macrophages (van Rossum et al., 2005), which aid the adaptive immune system to clear colonising bacteria from the mucosal surfaces (Martner et al., 2009). The adaptive immune response involves secretion of specific antibodies such as IgA and serum IgG, which are mainly produced by B cells (LeBien and Tedder, 2008). These antibodies mediate the clearance of pneumococcal strains colonising the nasopharynx (Joyce et al., 2009, Sun et al., 2004). A decrease in carriage rates from more than 50% in children to less than 10% in adults has been shown to correlate with increasing levels of both mucosal and serum antibodies to pneumococcal surface polysaccharides (Syrjanen et al., 2001, Zhang et al., 2004, Simell et al., 2002). The antibody response to pneumococcal proteins of carried strains was investigated in an experimental model of human colonisation in healthy adults in the USA (McCool et al., 2002). It was noted that all of the colonised subjects mounted a response to pneumococcal surface protein A (PspA) through secretion of serum IgG and secretory IgA, whereas subjects not colonised had pre-

existing antibody to this protein (McCool et al., 2002). A similar study in the UK (Wright et al., 2012) showed elevated levels of antigen-specific IgG and IgA in nasal washes 6 weeks after pneumococcal carriage challenge compared to baseline. These immunoglobulins were targeted towards several proteins on the surface of the pneumococcus and not the polysaccharide capsule (Wright et al., 2012). For example, the challenge with 23F increased IgG response towards PspA in bronchoalveolar lavage (BAL). Thus pneumococcal challenge immunised mucosal surfaces through production of anti-protein immunoglobulin (Wright et al., 2012). Antibodies are also thought to play an important role in the clearance of colonisation because of decreased rates of carriage of vaccine serotypes in vaccinated groups (Dagan et al., 2002, Dagan et al., 1997). Findings from other studies, however, seem to suggest the hypothesis that natural immunity to colonisation can be acquired independently of anti-capsular antibodies (McCool and Weiser, 2004). This hypothesis was tested using mutant mice with an impaired response to polysaccharide antigens (McCool and Weiser, 2004). It was observed that colonising strains were cleared at the same rate as the wild type (parent) strain. It was further demonstrated that mice, which lacked mature B cells and failed to produce antibodies, were unaffected in the density or duration of colonisation (McCool and Weiser, 2004). This clearly suggested that antibodies were not required for clearance of pneumococcal colonisation at least in a murine model (McCool and Weiser, 2004). It has also been shown that immunity to pneumococcal colonisation can be induced in the absence of antibody, however, the induced immunity requires the presence of CD4 T cells (Malley et al., 2005). The role of CD4 T cells was investigated in mice that were T cell-deficient, and those that were congenitally deficient or depleted of CD4 T cells at the time of the challenge. It was noted that

intranasal immunisation with whole-cell vaccine was completely abolished in these mice mutants (Malley et al., 2005). Another related study investigated the CD4 T cell proliferative response to pneumococcal virulence proteins in peripheral blood mononuclear cells (PBMCs) and adenoidal mononuclear cells (MNCs) taken from children (Zhang et al., 2007). It was noted that the proliferation of CD4 T cells after stimulation of PBMCs and MNCs with protein antigens was significantly lower in children who were culture positive for pneumococcus than in those who were culture negative, which suggested the role CD4 T cells play in modulating pneumococcal carriage in children (Zhang et al., 2007). An increase in CD4 T cell response has been reported to increase with age, which also correlates with a decrease in carriage (Cobey and Lipsitch, 2012) and pneumococcal disease, suggesting CD4 T cells play a role in controlling pneumococcal carriage and disease (Pido-Lopez et al., 2011, Malley et al., 2003). HIV infection is known to decrease the hosts CD4-cell levels (Le et al., 2013) and this might have an impact on pneumococcal carriage. Recently, an increase in pneumococcal colonisation with decrease in peripheral blood CD4 T-cell counts and progression of HIV infection in adults (Glennie et al., 2012a) has been reported. Further, the HIV and related inflammation are known to impair the epithelial barrier function and integrity leading to chronic bacterial colonisation. TH17 cells have been shown to mediate acquired immunity to pneumococcal colonisation in mice, which would also be expected to occur in humans (Lu et al., 2008), however, HIV infection impairs adaptive TH17 and memory T- and B-cell responses, which may fail to mediate the clearance of mucosal bacteria, or control bacterial multiplication following dissemination (Glennie et al., 2010).

1.5.3. Evading clearance by the host immune system

The host immune system employs among others, secretory IgA (sIgA) (Davis et al., 2008), lactoferrin (Shaper et al., 2004), and complement (Picard et al., 2003, Bogaert et al., 2010), to target the pneumococcus at the mucosal site. The sIgA prevents pneumococcal binding and promotes bacterial opsonisation, which enables phagocytosis by macrophages/monocytes and neutrophils (Finn et al., 2002). Naturally, the polysaccharide capsule of the pneumococcus acts as a physical barrier to opsonisation by sIgA (Fasching et al., 2007). The pneumococcus also releases IgA1 protease, which degrades the host's IgA (Weiser et al., 2003).

Lactoferrin on the other hand binds to iron, which is necessary for bacterial metabolism (Raphael et al., 1989). This interferes with the viability of the pneumococcus in the nasal cavity, consequently preventing colonisation. Free lactoferrin (apolactoferrin) has also been reported to have direct bacterial killing properties (Shaper et al., 2004, Arnold et al., 1980), thought to occur through bacterial cell lysis (Senkovich et al., 2007). The pneumococcus survives the killing effect of apolactoferrin by expressing a choline binding pneumococcal surface protein A (PspA), which inhibits apolactoferrin by binding to its active site (Shaper et al., 2004).

Complement plays an important role in the immune response against *S. pneumoniae* through bacterial opsonisation and phagocytosis (Bogaert et al., 2010, Mortensen and Duszkiewicz, 1977). To activate the complement pathway, the innate immune defence system utilises C-reactive protein, which binds to phosphorylcholine present on the pneumococcal cell surface (Gould and Weiser, 2001, Winkelstein and

Tomasz, 1977). To overcome complement dependent killing or clearance from the mucosal surfaces, the pneumococcus employs the polysaccharide capsule, which prevents complement deposition on pneumococcal cell surface (Hyams et al., 2010). Pneumococcal proteins such as pneumolysin (Ply) (Quin et al., 2007), pneumococcal surface protein C (PspC) (Dave et al., 2004a), PspA (Quin et al., 2007), and PhpA (Zhang et al., 2001) have also been shown to inhibit the deposition of complement on pneumococcal surface, consequently preventing complement mediated opsonisation and phagocytosis.

1.5.4. Pneumococcal binding to host epithelial cells

Once the host adherence barrier mechanisms are overcome, the pneumococcus makes some structural adjustments in readiness for binding to epithelial cells (Cundell et al., 1995b). Notably, the pneumococcal polysaccharide capsule undergoes phase variation from opaque (thick capsule) to transparent (thin capsule) phase, to expose binding sites on the cell wall surface (Cundell et al., 1995b, Weiser et al., 1994). An induction of phase variation is thought to occur during the approach of the pneumococcus to the epithelium. The pneumococcus binds to epithelial cell glycoconjugates such as N-acetyl-D-galactosamine (GalNac) (Figure 1.9A). The binding process is facilitated by the pneumococcal glycosidase Neuraminidase A (NanA), which increases the expression of binding sites on GalNac via sialic acid cleavage (King et al., 2004, Tong et al., 2001).

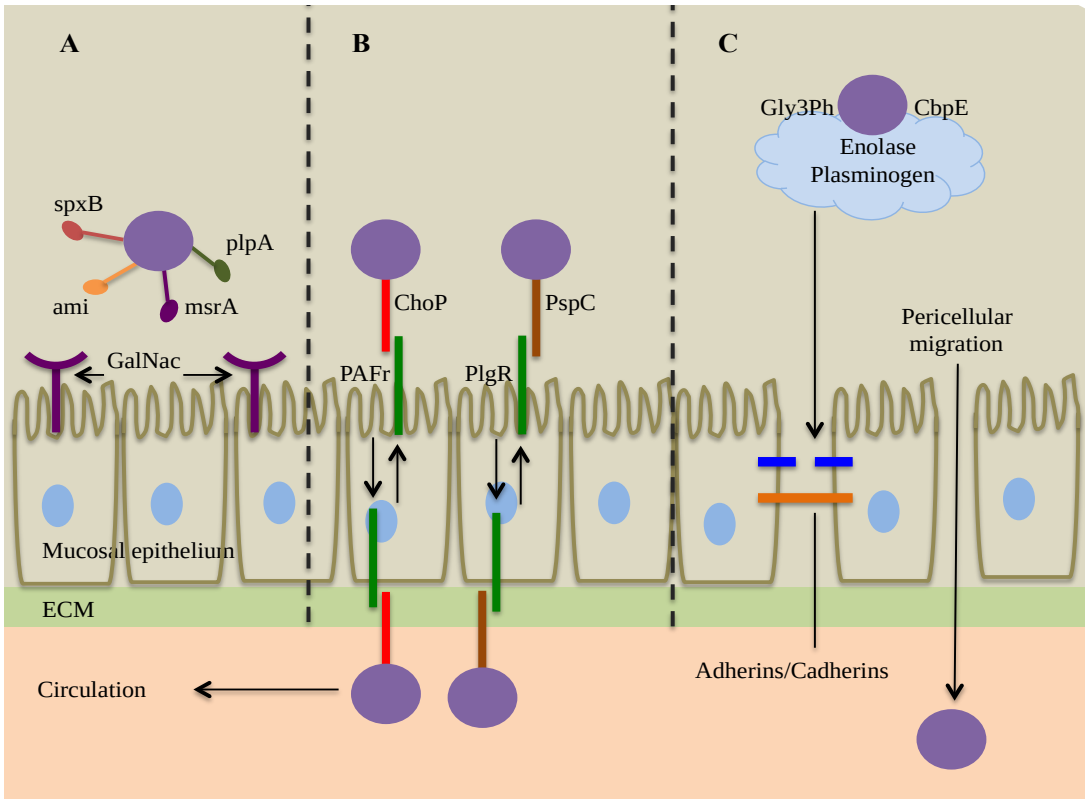


Figure 1.9 Pneumococcal binding to epithelial cells and internalisation

The pneumococcus initially binds to N-acetyl-D-galactosamine (GalNac) (A) followed by attachment to epithelium cell receptors PAFr and PIgR with pneumococcal surface proteins ChoP and PspC respectively (B), which facilitates internalization by the cell receptors recycling pathway. The pneumococcus also crosses the mucosal epithelium into circulation using plasminogen, which degrades proteins (adherins) holding cells together. This creates an open gap between the cells for migration of the pneumococcus into circulation (C) (Mook-Kanamori et al., 2011).

1.5.5. Pneumococcal translocation through epithelial cells into blood

Following the attachment of the pneumococcus to epithelial cells by GalNac, translocation into circulation is initiated by the binding of the pneumococcus to the host platelet activating factor receptor (PAFr) (Cundell et al., 1995a) and polymeric immunoglobulin (Jordano et al.) using phosphorylcholine (ChoP) and pneumococcal surface protein C (PspC) respectively (Figure 1.9B) (Cundell et al., 1995a, Elm et al., 2004). The pneumococcus is transported through the mucosal barrier into circulation by the recycling pathway of the PAF and pIgR receptors (Figure 1.9B). It has been reported that the ChoP-PAFr interaction is common among virulent strains of *S. pneumoniae* (Cundell et al., 1995a). The pneumococcus is also translocated into circulation using plasminogen (Pancholi et al., 2003), which is attached to the pneumococcal surface by enolase, Gly3Ph and CbpE (Figure 1.9C) (Attali et al., 2008b). Plasminogen has been reported to degrade protein molecules such as adherins, which play an important role in holding epithelial cell junctions together (Attali et al., 2008a). Once the junctions are broken, the pneumococcus is allowed easy passage across epithelial cells into circulation, consequently causing invasive disease.

1.6. Interventions for pneumococcal infections

1.6.1. Antibiotics

The most commonly used antibiotics in the fight against pneumococcal infections include β-lactam antibiotics such as penicillin and cephalosporins, macrolides, and trimethoprim-sulfamethoxazole (TMP-SMX). There is however increasing resistance in *S. pneumoniae* strains to these drugs worldwide (Adegbola et al., 2006, Marton et al., 1991, Spika et al., 1991, Karlowsky et al., 2003), including Malawi (Cornick et

al., 2011, Cornick et al., 2013, Everett et al., 2011). In general, resistance to antibiotics occurs through several biochemical mechanisms: mutations on the target protein/gene, drug inactivation through enzymatic activity, transfer of antibiotic resistance genes, or prevention of drug access to their targets (Nikaido, 2009). The mode of resistance in pneumococcus to β-lactam antibiotics is by alteration in the cell wall penicillin binding proteins (PBPs). PBPs play an important catalytic role in the synthesis of pneumococcal peptidoglycan. Beta-lactam antibiotics bind PBPs and reduce their ability to synthesise peptidoglycan, consequently affecting the integrity of the cell wall and viability (Zapun et al., 2008). Thus any alteration in PBPs of a given strain results in decreased affinity to β-lactam antibiotics. Point mutations in genes encoding PBPs have been implicated in development of pneumococcal resistance to β-lactam antibiotics (Reichmann et al., 1996, Hakenbeck et al., 1999, Laible et al., 1991). Macrolides function by inhibiting synthesis of bacterial proteins through binding to the 23S ribosomal RNA. Pneumococcal strains confer resistance to macrolides through modification of the target site mediated by *erm* genes (Edelstein, 2004) or efflux mechanism to remove the antibiotic from the cell by employing *emf* genes (Cochetti et al., 2005). TMP-SMX (co-trimoxazole) inhibits dihydrofolate reductase (DHFR) and dihydropteroate synthase (DHPS), which are involved in folic acid biosynthetic pathway, by presenting as false substrate inhibitors. Resistance to co-trimoxazole is developed when there are mutations occurring in genes encoding DHFR (*fol*A) and DHPS (*fol*P) enzymes, such as Ile-100-Leu mutation, together with a residue 92 substitution of DHFR (Cornick et al., 2013). Resistance to multiple antibiotics has also been observed. This is thought to occur by two possible mechanisms: (i) acquisition of an array of genes within a single bacterial cell, each of which confers resistance to a particular drug. Such

resistance genes are usually assembled in recombinant plasmids and/or transposons (Nikaido, 2009, Ding et al., 2009) and maintained within the bacterial cells within the host enabling their efficient transfers within the community; and (ii) increased expression of genes that code for multidrug efflux pumps, resulting in a wide range of drugs being pumped out of the cells (Nikaido, 2009).

1.6.2. Vaccination

Development of resistance to antibiotics globally, has adversely limited their usage in the control of pneumococcal infections, and this has made the need for vaccine usage even greater. The efficacy of pneumococcal vaccines is based on the production of opsonising antibodies, and currently, research has focused on the development of vaccines which can be effective in children and particular risk groups such as immune-compromised patients (Bogaert et al., 2004b).

1.6.2.1. Polysaccharide based vaccines

The first widely licensed vaccine to prevent pneumococcal infections was a 23-valent polysaccharide vaccine (PPV23) (WHO., 2008). PPV23 is composed of pure capsular polysaccharides from 23 invasive pneumococcal serotypes (1, 2, 3, 4, 5, 6B, 7F, 8, 9N, 9V, 10A, 11A, 12F, 14, 15B, 17F, 18C, 19F, 19A, 20, 22F, 23F, and 33F). These polysaccharides induce the production of anti-capsular immunoglobulin M (IgM) via a T-cell independent immunological response (Stein, 1992) (Figure 1.10). The response of B-cells is enhanced by the presence of cleavage products of complement factor 3 (C3), a ligand for complement receptor 2 (CR2) (Griffioen et al., 1991). This is more prominent in adults because young children do not effectively express CR2 (Griffioen et al., 1992). This may contribute to the poor

immunogenicity of PPV23 in children under 2 years of age (2008). PPV23 is a T-cell independent antigen and therefore fails to provide any significant protection against mucosal pneumococcal infections or against the spread of pneumococcal strains within the community (Eskola and Anttila, 1999). In Uganda, PPV23 was ineffective in HIV infected adults (French et al., 2000) and a very low efficacy was observed in HIV negative adults (French et al., 1998). In the United States, an alternative conjugate vaccine, the seven-valent pneumococcal conjugate vaccine (PCV7) was introduced into the routine infant immunisation programme in the United States in 2000 (Hausdorff et al., 2000). PCV7 has now been subsequently replaced with higher valent conjugate vaccines such as PCV10 and PCV13 (Table 1.3), to cover serotypes that emerged to cause invasive.

Table 1.3 Pneumococcal vaccine serotypes

Vaccine	Carrier Protein	Serotypes Covered												
7-valent	CRM_{197}	4	6B	9V	14	18C	19F	23F						
10-Valent	*H. Influenzae* proteinD, Tetanus and Diphtheria Toxoid	4	6B	9V	14	18C	19F	23F	1	5	7F			
13-Valent	CRM_{197}	4	6B	9V	14	18C	19F	23F	1	5	7F	3	6A	19A

A list of vaccine serotypes included in the 7-, 10- and 13-valent pneumococcal conjugate vaccines (adopted from (Isaacman et al., 2010)).

The pneumococcal conjugate vaccines are formulated by covalently attaching the capsular polysaccharide molecule to a non-toxic immunogenic peptide molecule. The polysaccharides in PCV13 for example are covalently bound to a common carrier protein, CRM^{197} (a nontoxic mutant diphtheria toxoid) (Esposito et al., 2010). The peptide-polysaccharide complex produced is presented in association with MHC II molecules to B-cells, thereby mounting a T-cell dependent immune response. The T-cell mediated immune response has the advantage of producing memory B-cells and enhancement of mucosal immune responses through the production of secretory IgA and IgG (Lottenbach et al., 1999, Nurkka et al., 2001, Simell et al., 2012b). How polysaccharide and conjugate vaccine activate the immune system is summarised in the figure below (Figure 1.10) (Ada, 2001).With polysaccharide vaccination, the polysaccharide antigen is bound to the surface of a B-cell in lymphoid tissues. The immune response to polysaccharide antigens is characterised by the production of IgM antibodies (Figure 1.10A), isotype restriction and delayed development (Weintraub, 2003). Conjugate vaccines on the other hand can bind in two ways (Figure 1.10 B): (i) they bind to dendritic cells by presenting the protein moiety of the vaccine to type 2 helper T (Th2) cells; or (ii) they bind to B-cells which express surface IgM receptors specific for the carbohydrate component of the polysaccharide-protein conjugate vaccine. Once bound, endocytosis and processing by the B-cells occurs to produce peptides, which are expressed with class II MHC molecules on the surface of B-cells to form the MHC-peptide complex. The MHC-peptide complex is bound by the activated Th2 cell and triggers the release of cytokines such as interleukin (IL)-4, IL-5 and IL-6, which cause the B-cell to undergo differentiation and express IgG molecules, which are specific for the polysaccharide component of the vaccine. These cells are reported to mature in

lymphoid follicles and cells expressing IgG antibodies become plasma cells (Ada, 2001).

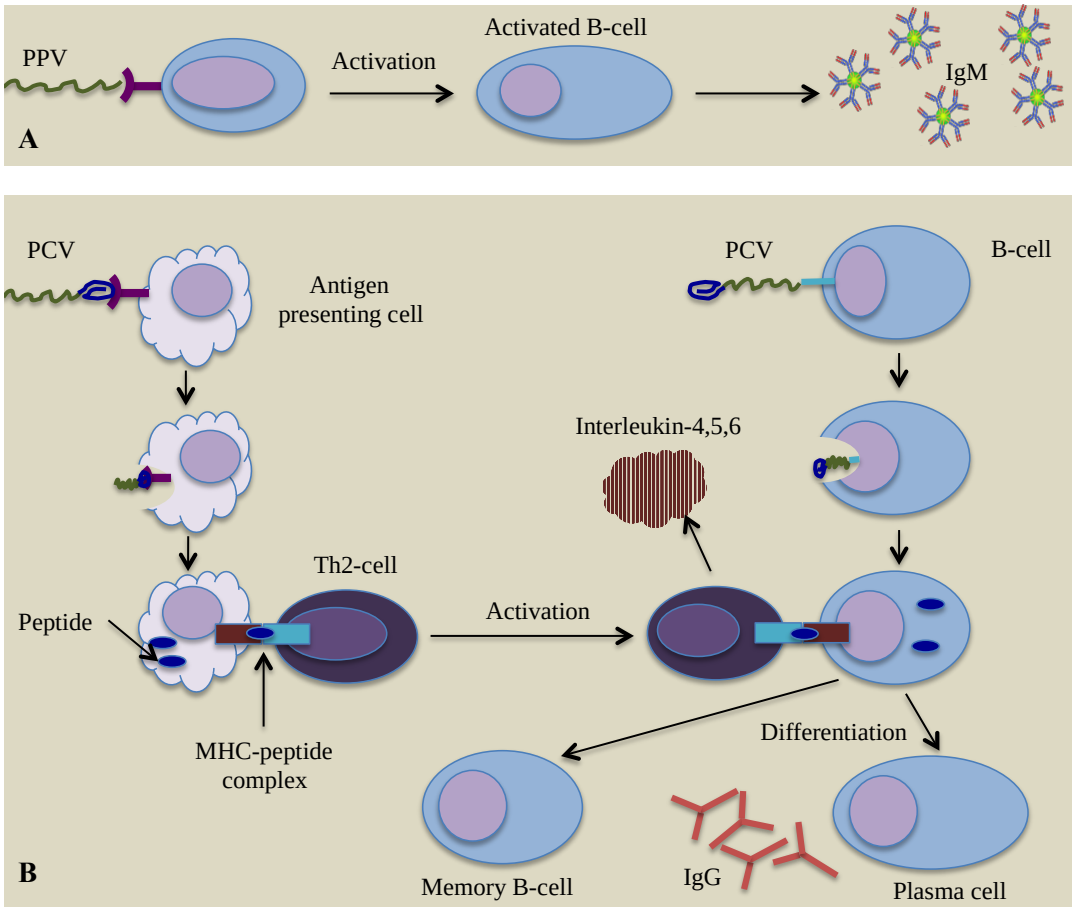


Figure 1.10 Antibody response to PPV (A) and PCV antigens (B)

Covalent coupling of PPV antigen with a protein carrier changes the antigen to T cell-dependent (TD). The antigen-presenting cells take up the conjugated PCV molecule, internalize it via the membrane immunoglobulin and present the peptides of the protein to the T helper cells in association with the major histocompatibility complex Class II molecules on their surface. This induces the T helper cells to stimulate PS-specific B cells to mature either into antibody-producing plasma cells or into memory cells (Ada, 2001).

Unlike PPV23, vaccination with conjugate vaccines has been shown to reduce the rates of asymptomatic carriage of vaccine serotypes (Pletz et al., 2008), and also demonstrated clinical efficacy against IPD caused by vaccine serotypes (VT) in infants and young children (Black et al., 2000, Poehling et al., 2006), including immuno-compromised HIV-infected adults (French et al., 2010). The USA has seen a nearly 100% reduction of invasive disease caused by vaccine serotypes in children under the age of 5 years (Pilishvili et al., 2010). Conjugate vaccination has also been shown to prevent carriage of serotypes covered by the vaccine (Roca et al., 2011, O'Brien et al., 2007), which prevents the spread of such serotypes in a community thereby ensuring protection in the unvaccinated groups (herd immunity) (Isaacman et al., 2007, Hammitt et al., 2006). The potential limitation facing conjugate vaccines is serotype replacement in both disease and carriage (Miller et al., 2011). In the United States, a 140% increase in the rate of invasive pneumococcal disease from non-vaccine serotypes has been reported in children less than 2 years old in the post PCV7 era (Singleton et al., 2007). Serotype replacement has also been reported in resource poor settings such as Gambia, where carriage of non-vaccine serotypes was high (76.9 %) in children who had received three doses of pneumococcal conjugate vaccine, compared with 42.5% in controls (Obaro et al., 1996). Similarly, in South Africa, in trials of a 9-valent vaccine, there was an increase in carriage of non-vaccine serotypes from 21 % in controls to 39 % in vaccine recipients (Mbelle et al., 1999). Conjugate vaccines with wider serotype coverage such as PCV10 and PCV13 (Dagan, 2009) have taken into account serotypes such as 1, 5, 7F, 3, 6A and 19A (Table 1.3), known to cause invasive disease in the post PCV7 era (Isaacman et al., 2010). In February 2010, the USA added PCV13 to its routine immunization program for infants and young children (Li et al., 2009). Worldwide, 96 countries

have incorporated PCVs into routine childhood immunisation programs and 51 countries in the developing world plan to introduce PCV in the near future (VIMS, 2013). Through the Global Alliance for Vaccines and Immunisation (GAVI) program, the Malawian government introduced PCV13 into the routine infant immunisation schedule in November, 2011 (Everett et al., 2012). The estimated coverage range of PCV13 in Malawian children was 70-90 % (Everett et al., 2012, Cornick et al., 2011). However, it is likely that PCV10 and PCV13 will face similar limitations of serotype replacement (Mulholland and Satzke, 2012) associated with previous capsular-based vaccines.

1.6.2.2. Protein based vaccines

To address the shortcomings by polysaccharide based vaccines, pneumococcal proteins are currently being explored as targets for development of effective pneumococcal vaccines (Ogunniyi et al., 2007a, Darrieux et al., 2013), with several protein candidates at various stages of clinical trials (Table 1.4) (Darrieux et al., 2013). It is thought that vaccines based on invariable and conserved proteins will provide cheap, non-serotype dependent protection without compromising efficacy. Protein based vaccines are expected to be highly immunogenic because they are T-cell-dependent antigens, and therefore likely to elicit immunological memory in children under the age of 2 years (Ogunniyi et al., 2007a). Previous studies reported that monoclonal antibodies against pneumococcal surface protein A (PspA) were able to protect mice from invasive pneumococcal disease, even when challenged with pneumococcal strains of more than one serotype (McDaniel et al., 1991). Although PspA is common to all pneumococcal strains, it is divided into three families based on nucleotide and amino acid identity, and each family is further

subdivided into different clades. Emerging data seem to suggest that PspA family and clade distribution are independent of serotype, age and clinical origin of the strains, but are strongly associated with genotype (Rolo et al., 2009). Thus while considering proteins as vaccine targets, it is important to pay attention to genetic variations that might exist within different pneumococcal proteins. These variations may arise from antigen diversity due to gene presence or absence, gene sequence variation (e.g. *pspA* gene) and differential expression. A novel surface protein, pneumococcal collagen like protein A (PclA) reported recently (Paterson et al., 2008) is a good example of how protein antigen diversity might affect the coverage of protein based vaccines. This protein has been shown to contribute to adherence and invasion of host cells by *S. pneumoniaein vitro*, but it demonstrated selective strain distribution (Paterson et al., 2008). The *pclA* gene is not present in all pneumococcal strains and this selective distribution was also observed in Malawian strains (Croucher et al., unpublished data). It has also been observed that although some virulent pneumococcal proteins elicit a significant level of protection in animal models, a large number of them remain uncharacterised, and to date, no single protein has been able to elicit protection comparable to that achieved by pneumococcal conjugate vaccines (Ogunniyi et al., 2007a). Thus exploring the possibility of protein antigens in combination with other proteins or with polysaccharides may provide more complete and effective coverage than is currently the case. A synergistic effect has been observed in animal models using combinations of PdB, PspA and PspC even in conditions where individual proteins offered little protection (Ogunniyi et al., 2007a).

Table 1.4 Pneumococcal protein vaccine targets

A

Candidate protein	Function	Immunological response
PspA* (pneumococcal surface protein A)	Inhibits complement activation (Mukerji et al., 2012, Ren et al., 2012) Inhibits killing by lactoferrin peptides (Mirza et al., 2011, Shaper et al., 2004)	Induces antibodies that recognise surface of pneumococcal strains and promote opsonisation by C3 deposition (Ochs et al., 2008, Ren et al., 2012)
NanA and NanB Neuraminidases	Promotes colonisation (NanA) and survival in blood (NanB) (Manco et al., 2006) by cleaving terminal sialic acid residues from glycoconjugates (King et al., 2006b)	Antibodies against NanA reduces risk of Otitis media (Long et al., 2004)
PiuA and PiaA	Components of the ABC transport system (Kadioglu et al., 2008) required for full virulence in animal models of pneumococcal infection (Brown et al., 2001a)	Antibodies raised against these two proteins were protective against invasive disease by *S. pneumoniae*(Brown et al., 2001b)

*The proteins are listed in sub-tables A and B. *Proteins that have gone through phase I or II of clinical trials (Darrieux et al., 2013)*

B

Candidate protein	Function	Immunological response
PspC ***Pneumococcal surface protein C***	Binds secretory IgA, prevents C3 deposition and mediates translocation to sterile sites such as blood (Dave et al., 2004b, Dave et al., 2004c)	Antibodies raised against PspC protective against colonisation (Balachandran et al., 2002, Hernani Mde et al., 2011). Protection against systemic infection inconclusive (Ferreira et al., 2009, Ogunniyi et al., 2001)
PsaA* ***Pneumococcal surface antigen A***	Plays a role in manganese transport (Dintilhac et al., 1997) resistance to oxidative stress (Tseng et al., 2002)and bacterial adhesion (Berry and Paton, 1996, McAllister et al., 2004)	Antibodies raised against PsaA protective against carriage (Pimenta et al., 2006, Miyaji et al., 2001, Oliveira et al., 2006), with minimal effect against systemic disease (Talkington et al., 1996, Gor et al., 2005)
Ply ***Pneumolysin***	Cholesterol dependent pore forming cytolysin (Kadioglu et al., 2008)	Ply recognition by TLR-4 activates innate immune response to pneumococcal infection(Malley et al., 2003)
PcpA * Pneumococcal choline-binding protein A	Possible role in bacterial adhesion (Seepersaud et al., 2005) and is expressed during invasive disease in lungs and blood and not nasal mucosa (Glover et al., 2008, Johnston et al., 2006)	Antibodies against PcpA protective against invasive disease (Glover et al., 2008) and colonisation (Khan and Pichichero, 2012)

C

Candidate protein	Function	Immunological response
PhT family *(PhTA,PhTB, PhTC, PhTD*, PhTE)*	Inhibit complement deposition on the bacterial surface and to bind zinc(Ogunniyi et al., 2009, Rioux et al., 2011)	PhT proteins were shown to confer protection in animal models of nasal and lung colonisation, lethal intranasal challenge and sepsis. PhTD was shown to be the most protective(Adamou et al., 2001, Godfroid et al., 2011)
Whole cell pneumococcal vaccine (WCV)*	Shown to induce IL-17 mediated protection against intranasal challenge, induces antibody formation which is independent of CD4+ T cells(Lu et al., 2010)	Protects against intranasal colonisation by multiple pneumococcal serotypes(Malley and Anderson, 2012)
Ply *Pneumolysin*	Cholesterol dependent pore forming cytolysin(Kadioglu et al., 2008)	Ply recognition by TLR-4 activates innate immune response to pneumococcal infection(Malley et al., 2003)
PcpA * Pneumococcal choline-binding protein A	Possible role in bacterial adhesion (Seepersaud et al., 2005) and is expressed during invasive disease in lungs and blood and not nasal mucosa (Glover et al., 2008, Johnston et al., 2006)	Antibodies against PcpA protective against invasive disease (Glover et al., 2008) and colonisation (Khan and Pichichero, 2012)

1.7. Methods for characterising *S. pneumoniae* in carriage

1.7.1. Phenotypic methods

The World Health Organisation established a working group on pneumococcal carriage to formulate standard methods for assessing pneumococcal serotypes for purposes of PCV vaccine efficacy studies (O'Brien and Nohynek, 2003). The initial step in pneumococcal characterisation involves positively identifying the pneumococcal strains from carriage based on their sensitivity to Optochin (Bowers and Jeffries, 1955) (Figure 1.11). For strains that are optochin resistant, a further bile solubility test is conducted (Murray, 1979); where a positive identification is by the solubility of the isolate; whereas insolubility or aggregation shows the presence of nontypeablepneumococcal strains, acapsular pneumococcal strains or non-pneumococcal strains. The positively identified pneumococcal strains are then further analysed to identify the serotypes present (Satzke et al., 2014).

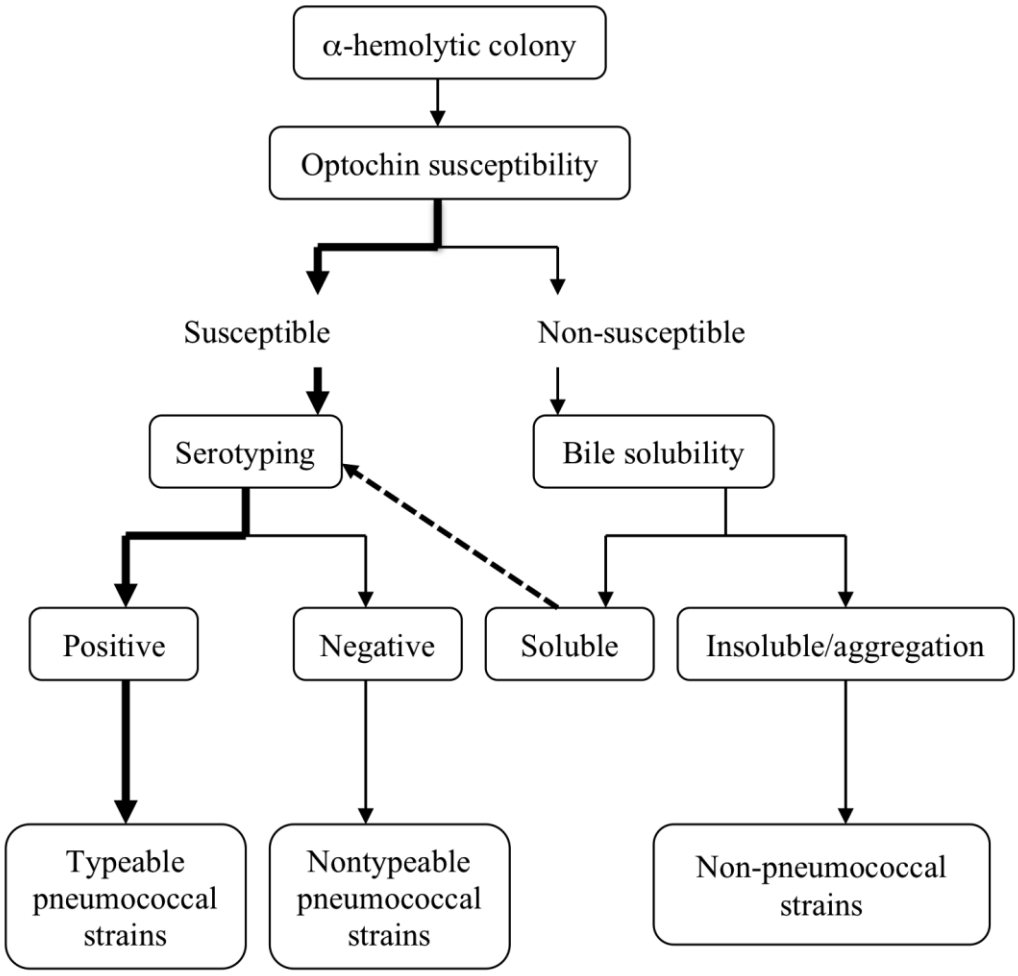


Figure 1.11 Pneumococcal detection method in carriage

The bold line shows the normal route of characterising optochin sensitive colonies to serotype pneumococcal strains. Optochin resistant colonies are characterised by bile solubility testing. Bile soluble colonies indicate presence of pneumococcal strains and these are further characterised by serotyping (dashed arrow line) (Satzke et al., 2014).

Quellung reaction. Quellung reaction is the current gold standard for serotyping pneumococcal strains as recommended by the WHO pneumococcal carriage working group (Satzke et al., 2014). Quellung involves the capsular reaction test

(Quellung/Neufed test) with antisera against the 95 known pneumococcal polysaccharide capsules (O'Brien and Nohynek, 2003, Satzke et al., 2014). A positive result for Quellung reaction involves binding between the capsular polysaccharide of pneumococcal strains and the type specific antibody present in the typing sera. This binding causes a change in the refractive index of the capsule so that it appears swollen, and more visible under a light microscope (http://www.cdc.gov). Serotyping by Quellung initially starts with pooled antisera in succession until a positive result is seen. To determine sero-group and serotype, testing with individual group and serotype-specific antisera included in the pooled antisera is subsequently performed. The sensitivity of the Quellung reaction has been validated in several laboratories across Europe (Konradsen, 2005). Although Quellung serotyping is preferred in laboratories with qualified personnel and enough resources, it is still a costly method to perform, is time consuming and requires high level of experience to perform satisfactorily (Satzke et al., 2014, Slotved et al., 2004, Morais et al., 2007).

Latex agglutination. The WHO pneumococcal carriage-working group has also recommended usage of latex agglutination for characterising pneumococcal strains. Latex agglutination uses anti-rabbit IgG-coated latex particles sensitized to pooled and selected individual serotype-specific antisera (http://www.cdc.gov). Latex agglutination is known for its sensitivity and specificity (Satzke et al., 2014). Compared with Quellung, the latex agglutination method is less expensive, easier to use and does not require a microscope (Gella et al., 1991, Satzke et al., 2014, Slotved et al., 2004). The disadvantage of latex agglutination is that it is prone to cross-reactions, which result in non-specific binding, consequently giving rise to false

positive results (Ingram et al., 1979). Latex agglutination methods, including the Quellung reaction have been shown to underestimate the prevalence of multiple carriage (Turner et al., 2011). In addition, these methods cannot always differentiate *Streptococcus mitis*, *Streptococcus oralis*, or nontypeable pneumococcal strains from typeable pneumococcal strains (Arbique et al., 2004b, McEllistrem, 2009).

1.7.2. Genotypic methods

Current genotypic methods are based on the knowledge of the pneumococcal genome. The sequenced genome of *S. pneumoniae* consists of a single circular chromosome of approximately 2 million base pairs, a G+C content of 40 % and more than 2000 open reading frames (ORFs) (Croucher et al., 2009, Tettelin et al., 2001). The serotype of a given pneumococcal strain is determined by its polysaccharide capsule. The region on *S. pneumoniae* genome, responsible for the synthesis of capsular polysaccharides (CPS locus) has been well characterised. The CPS locus is located between *dexB* and *aliA* genes (Figure 1.12) and the total size of the alternative coding DNA at this locus is approximately 1.8Mbp (Bentley et al., 2006).

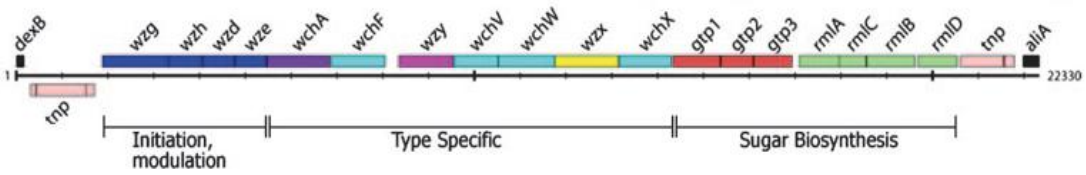


Figure 1.12 A representative *CPS* locus for serotype 23F

Capsule biosynthesis genes are shown between *dexB* and *aliA* genes (Aanensen et al., 2007).

The capsular polysaccharides (CPSs) are synthesised by two pathways namely: the Wzx/Wzy-dependent pathway and the synthase pathway. The CPS genes of 93 pneumococcal serotypes are synthesised by the Wzx/Wzy pathway, while serotypes 3 and 37 utilise the synthase pathway, encoded by *wchE* and *tts* genes respectively (Arrecubieta et al., 1994, Bentley et al., 2006, Llull et al., 1999). The genes encoding enzymes for the Wzx/Wzy and synthase pathways are located at the same chromosomal location between *dexB* and *aliA* genes, except *tts* gene for serotype 37, which is located elsewhere on the pneumococcal chromosome (Llull et al., 1999). The CPS biosynthetic genes are known to be orientated in the same direction as the *dexB* and *aliA* genes (Bentley et al., 2006). The four genes (*wzg, wzh, wzd, wze*) at the 5' end of the CPS locus are involved in initiation and modulation, and they are conserved with high sequence identity in all strains (Bentley et al., 2006). The fifth gene, *wchA*, encodes the initial glucose phosphate transferase (WchA) in most CPSclusters. *WchA* is responsible for the linkage of glucose phosphate to a membrane – associated lipid carrier, initiating the repeat unit biosynthesis (Aanensen et al., 2007, Bentley et al., 2006). Among the genes always present downstream are the flippase (*wzx*) and polymerase (*wzy*) genes, which play a role in transporting the repeat unit across the cytoplasmic membrane and linking individual repeat units to form lipid-linked CPS (Bentley et al., 2006). Unlike the four genes upstream, there are multiple highly divergent or nonhomologous groups of genes such as *wzx* and *wzy*, which might contribute to the observed variations in the polysaccharide capsule structure in various pneumococcal serotypes (Mavroidi et al., 2007).

1.7.2.1. PCR based methods

PCR based methods have widely been used to identify pneumococcal serotypes in clinical samples. These methods include Multiplex PCR (Pai et al., 2006, Brito et al., 2003), Real-time PCR (Pimenta et al., 2013) and single PCR with sequencing (Leung et al., 2012). PCR serotyping is dependent on the knowledge of the *CPS* locus gene sequences (Bentley et al., 2006) in order to identify serotypes in a given sample. The basis of classification is the presence of genes, which are serogroup/serotype specific. The advantage of PCR methods is that they are highly sensitive, easy to use, and have the ability to detect non-viable organisms (Satzke et al., 2014). PCR techniques are also widely used in resource poor settings (Morais et al., 2007). Despite these advantages, PCR methods are not able to discriminate closely related serotypes, which are often detected as a group (Morais et al., 2007). In particular, there is risk of amplicon contamination for Multiplex PCR (Satzke et al., 2014).

1.7.2.2. Microarray based methods

Microarrays were developed following the availability of microbial whole genome sequences (Doolittle, 2002). Microarrays allow for the simultaneous analysis of every gene in the genome, either at the DNA or mRNA level (Ye et al., 2001). Microarrays are constructed by depositing probes such as nucleic acids, proteins, carbohydrates, and antibodies on a solid surface such as glass. There are two types of commonly used DNA microarrays namely: PCR product - based DNA microarray and oligonucleotide-based DNA microarrays. PCR product-based microarrays are prepared by spotting the PCR amplicons onto a glass microscope slide. Although PCR based microarrays offer advantages such as ease of production and relative low cost, they are prone to cross-hybridisation between genes that are very similar (Wren

et al., 2002), which makes it difficult to detect genetic insertions, inversions, duplications or indeed point mutations (Dorrell et al., 2002). Unlike PCR arrays, oligonucleotide – based DNA arrays are generated by identifying areas of specificity within any gene sequence as targets for hybridisation. Knowledge of the sequence of the target DNA is required for synthesising complementary oligonucleotides. The advantages of oligonucleotide over PCR based microarrays include a reduction in cross-hybridisation and an increase in distinguishing homologous genes (Dorrell et al., 2002). To determine the presence of target genes, nucleic acid samples are labelled, either chemically or by an enzymatic reaction and then hybridized onto the array (Figure 1.13). Unbound and non-specifically bound samples are removed by washing using different stringency buffers. The remaining signal resulting from specific interactions between probes and target nucleic acids is measured. Only probes that hybridize to a labelled complementary target will yield a signal, thereby ascertaining the presence of the gene of interest in the sample (Huyghe et al., 2009).

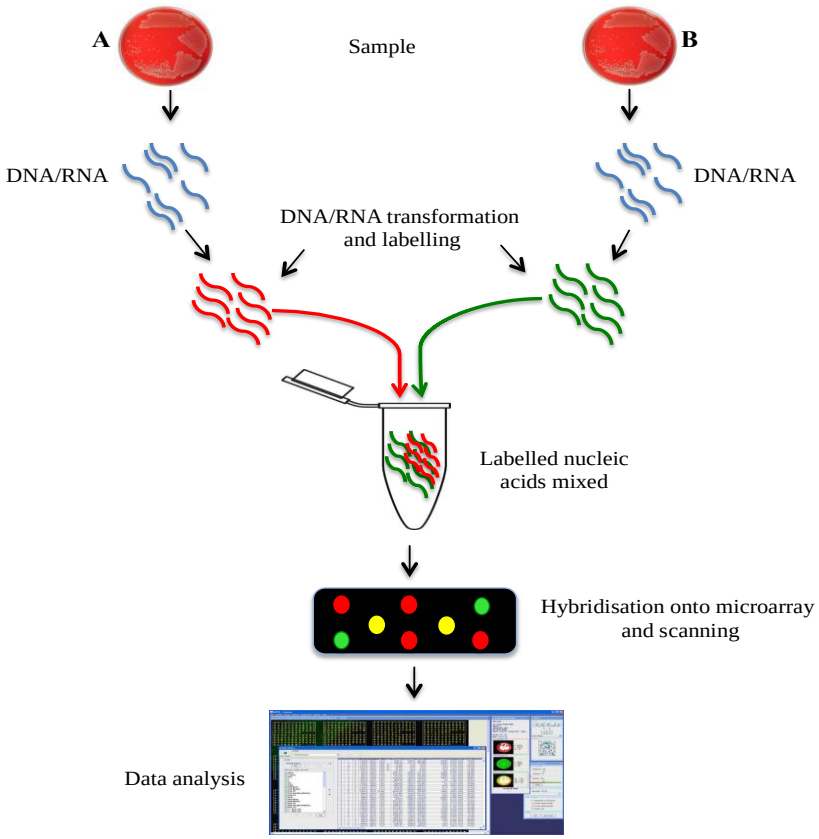


Figure 1.13 Microarray workflow

(http://grf.lshtm.ac.uk/microarrayoverview.htm)

Microarray technology has found wide applications in microbiology such as detection of differential gene expression, gene regulation as well as comparative genomics (Behr et al., 1999; Richmond et al., 1999; de Saizieu et al., 2000). However, the limitation of microarray analyses is that they are restricted to the genes present in the reference strain(s) on the microarray, which renders rapid identification of acquired DNA in outbreak strains difficult. This thesis describes studies using oligonucleotide microarray to characterise serotypes in carriage and also estimate the prevalence of multiple carriage in Malawian adults and children. Unlike Quellung or latex agglutination, microarray has the enhanced utility to detect

not only carriage of multiple strains of *S. pneumoniae* and their relative abundance but also detect the presence of antibiotic resistance genes, novel serotypes and other colonising microbial species in the nasopharynx (Newton et al., 2011, Turner et al., 2011). Microarray also offers the advantage of analysing large number of samples within a short period of time and can be optimised to detect non-viable organisms (Satzke et al., 2014).

1.7.2.2.1. *S. pneumoniae* capsular polysaccharide (SP-CPS) microarray

The SP-CPS microarray is a novel molecular based serotyping technique developed by the bacterial microarray group at St George's (BμG@S) in London. It was designed based on the availability of DNA sequences for all the genes responsible for the synthesis of the capsular polysaccharide (CPS), which determines serotypes. The BμG@S SP-CPSmicroarray (http://www.bugs.sgul.ac.uk) contained several thousands of 60mer oligonucleotides probes, with 10 oligonucleotide probes per gene (Figure 1.14).

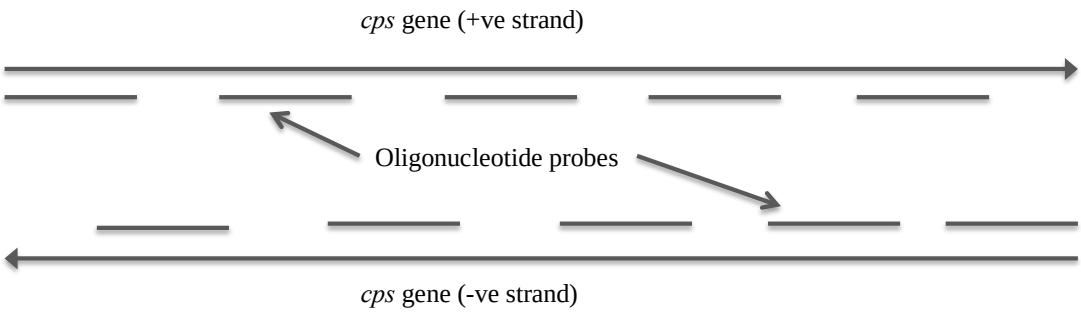


Figure 1.14 The CPSgene target regions of the 60-mer oligonucleotides

There are 10 oligonucleotides per CPS gene and these are spread along the length of the target CPS gene on both negative and positive strands and are randomly printed on the array. The use of multiple oligonucleotides per CPS gene plays an important role in minimising false positives due to cross hybridisation.

The long 60mer, in situ synthesised oligonucleotides DNA microarrays are known for their high sensitivity. They are also very tolerant of sequence mismatches and are thus suitable for the analysis of regions, which are highly polymorphic (Fenart et al., 2013). These probes detect a number of different entities (Hinds et al., 2009): (a) homology group identification (HGID) probes (Chapter 2, section 2.6.4.1) for determining serotype by the presence of capsular polysaccharide locus genes, with 10 oligonucleotide reporters per gene. There are 432 CPS genes known to date, however a particular serotype can only contain a subset of these genes (Newton et al., 2011); (b) serotype identification (STID) (Chapter 2, section 2.6.4.1) oligonucleotides designed to discriminate serotypes with identical HGID; (c) pathogen identification (PathID) probes (Chapter 2, section 2.6.4.1) for identifying bacterial species other than *S. pneumoniae*, that are commonly found in asymptomatic carriage; (d) antibiotic resistance (AbR) reporters (Chapter 2, section 2.6.4.1.1) for detecting the presence of antibiotic resistance genetic markers. The array used in this study was only designed to detect 10 antibiotic resistance genetic markers such as *aphA3* (kanamycin resistance) (Werner et al., 2001), chloramphenicol acetyltransferase (*cat)* gene (chloramphenicol resistance), erythromycin methylase (*ermB/C*) genes (erythromycin resistance), macrolide efflux (*mefA*) gene (macrolide resistance) (Arpin et al., 1999, Ardanuy et al., 2005), streptothricin acetyltransferase (*sat4)* gene(streptothricin resistance) (Werner et al., 2001), and *tetK/L/M/O* genes as markers for tetracyclines and macrolides resistance (Marimon et al., 2006). Apart from detecting serotypes, the BμG@S SP-CPS microarray also offers enhanced utility for novel serotype discovery; nontypeable (NT) strains investigation, detection of multiple serotype carriage (Chapter 2, section 2.6.4.3) and performing comparative genomic hybridisation (arrayCGH) using

spTIGR4+R6 genome backbone (Hinds et al., 2009). The sensitivity of the fluorescent detection enables serotypes in the order of 1% relative abundance to be detected in multiple carriage samples. The discovery of novel serotypes is achieved by detecting HG profiles that do not match the CPSgene content of any of the currently known reference strain serotypes (Hinds et al., 2009).

1.8. The human microbiota

The pneumococcus colonise the nasopharynx with a wide range of other microbiota, and interact constantly (Bosch et al., 2013). Such interactions have potential implications for microbial virulence through the exchange of genetic material and the aetiology of polymicrobial infections (Rodrigues et al., 2013b, Hiller et al., 2010). But the nasopharynx only contains a small part of the entire human microbiota, which will be briefly described here. From birth, the human body consists of only eukaryotic human cells. However, during the first years of life, the human skin surface (Kong et al., 2012), oral cavity (Takeshita et al., 2012), and gut (Yatsunenko et al., 2012) are colonised by a wide variety of microbial species such as bacteria, archaea, fungi, and viruses. Together, these microbial species constitute what is known as the human microbiome or microbiota (Consortium., 2012). The human microbiota contains ten times as many cells as the rest of the human body and orders of magnitude more genes than the human genome (Consortium., 2012, Qin et al., 2010, Whitman et al., 1998). The microbial species colonizing the human body are normally commensal and play beneficial roles such as food digestion as well as maintenance of the human immune system. However, their influence in human health and disease is still not fully understood (Huang and Lynch, 2011, Huang et al., 2011, Kong et al., 2012, Marchesi et al., 2011). Recent studies suggest an association

between disease and carriage of certain microbial species. In the USA a shift in the composition of rectal mucosal microbiota was associated with HIV infection. In this study, genera such as *Roseburia, Coprococcus, Ruminococcus, Eubacterium, Alistipes* and *Lachnospira* were depleted in HIV-infected subjects, followed by a simultaneous and significant enrichment of *Fusobacteria*, *Anaerococcus*, *Peptostreptococcus* and *Porphyromonas*. This change was also accompanied by a change in biochemical functions such as amino acid metabolism and vitamin biosynthesis, which differed significantly between HIV positive and HIV negative controls (McHardy et al., 2013). A shift in nasopharyngeal microbiota during acute otitis media characterised by a decrease in commensal bacteria such as *Staphylococcaceae*, *Flavobacteriaceae*, *Carnobacteriaceae* and *Comamonadaceae*, has also been reported in children in Switzerland (Hilty et al., 2012). Recently in the USA, a study showed a significant shift in the nature of gut microbiome during progressive HIV infection (Vujkovic-Cvijin et al., 2013). The authors reported an imbalance in the gut microbial community, which was characterised by a low representation of beneficial bacteroidia bacteria, and enrichment of proteobacteria in HIV-infected subjects. There was also an increase in bacteria responsible for catabolizing tryptophan through the kynurenine pathway, which correlated with kynurenine levels in HIV-infected subjects (Vujkovic-Cvijin et al., 2013). In Malawi, changes in the gut microbiome in children were implicated as a possible causal factor in kwashiorkor (Smith et al., 2013). These findings suggest that knowledge of the human microbiota and any deviation from the normal microbiota can potentially be used as an early detection biomarker for various disease conditions and as a target for therapeutic intervention (Morgan and Huttenhower, 2012).

1.8.1. Nasopharyngeal microbiota and upper respiratory tract infections

The bulk of microorganisms in the nasopharynx are commensals while a few may be pathogenic because they have been shown to cause disease. Comparing the microbial profile between subjects with a particular illness and healthy controls has revealed the potential role these pathogens play in disease (Hilty et al., 2012). A study in children in the Netherlands showed that the nasopharyngeal microbial profile was seasonal and different between samples collected from children in fall/winter and spring. It was further reported that Infections caused by these nasopharyngeal microbiota occurred primarily in fall and winter, however, it is not clear whether this was associated with the corresponding changes in the profile of nasopharyngeal microbiota (Bogaert et al., 2011). These studies however, were performed on HIV negative subjects. The data on the impact of HIV infection on human microbiome are limited particularly those relating to the nasopharynx. As previously described (Section 1.8), studies focusing on other parts of the body have shown a difference in the microbial communities between HIV positive and HIV negative subjects. It is clear from these findings that microbial patterns change during disease conditions, however it is not known how the nasopharyngeal microbial communities change during HIV infection, highlighting the need for further investigation particularly in settings with high HIV disease burden such as Malawi.

1.9. Microbial interaction in the nasopharynx

1.9.1. Bacterial-bacterial interaction in the nasopharynx

The interactions among bacterial species occupying the same niche have been well documented. These interactions could lead to coexistence (positive association) of several species in the same environment or suppression (negative association) of one

microbial species by another (Pettigrew et al., 2008), and have a major influence on which bacterial species persist in the nasopharynx (Jacoby et al., 2007, Pettigrew et al., 2008). To understand the mode of bacterial interactions in the nasopharynx, several mechanisms have been proposed. Coexistence for example is thought to occur through the release of components such as outer membrane vesicles containing proteins that may function to inactivate complement factor C3, thereby enabling another species to escape the immune system. *M. catarrhalis* has been shown to protect *H. influenzae* from complement-mediated killing through the release of protective outer membrane vesicles (Tan et al., 2007). Bacteria are also known to associate positively by growing and persisting in microbial communities known as biofilms, which is a universal attribute of bacteria (Lopez et al., 2010). Biofilms are dense aggregates of bacterial cells embedded in a self-produced extracellular matrix composed of polysaccharides, DNA and proteins which act as adhesins (Cvitkovitch et al., 2003, Lopez et al., 2010). Biofilm formation is important for bacterial survival as it enables bacteria to evade the host immune system (Domenech et al., 2013). Clinically, biofilm formation has been implicated in various infections such as urinary tract infections, chronic otitis media, and chronic tonsillitis (Hatt and Rather, 2008, Post et al., 2004, Hall-Stoodley et al., 2004) and may confer on the bacteria involved the ability to resist antibiotics resulting in infections that may not be easily treated (Ito et al., 2009, Mah and O'Toole, 2001). A recent study has also shown that biofilm formation plays an important role in genetic recombination by providing an enabling environment for species genetic transfer to occur (Marks et al., 2012).

Suppression on the other hand is characterised by the release of substances such as hydrogen peroxide (H_2O_2) by one species that are toxic to competing species. *S.*

pneumoniae is known to produce very high levels of H_2O_2, which kill *S. aureus* (Regev-Yochay et al., 2006) and *H. influenzae* (Pericone et al., 2000). *S. pneumoniae* is also reported to express neuraminidase enzyme, which cleaves sialic acids on the surface of *H. influenzae*, limiting its ability to colonise the mucosal surfaces (Shakhnovich et al., 2002). Bacterial species such as *S. pneumoniae* and *H. influenzae* express phosphorylcholine required for bacterial adherence to host mucosal surfaces. Pre-exposure to either one of the two species enhances the production of antibodies against phosphorylcholine, resulting in the elimination of other competing species (Tanaka et al., 2007, Goldenberg et al., 2004). It has also been shown in vitro, that phosphorylcholine is required for the survival of *S. pneumoniae* and not *H. influenzae*. To escape the host immune system, *H. influenzae* may switch off the expression of phosphorylcholine, which gives it a better chance of survival than *S. pneumoniae* (Goldenberg et al., 2004, Weiser et al., 1997). The interaction of microbial species with the host's immune system also plays a central role in the interspecies competition (Lysenko et al., 2005) and must be taken into account. For example, an epidemiological study in children found a significant negative association between *S. pneumoniae* and *S. aureus* among HIV negative carriers. This association was not observed in the HIV positives (Madhi et al., 2007b).

1.9.2. Bacterial and viral interaction in the nasopharynx

Bacteria are also in constant contact with other microorganisms present in the nasopharynx such as viruses. The nature of the interaction between bacteria and viruses has been well documented. One example of viral-bacterial interaction involves the positive association between influenza virus and *S. pneumoniae*, where

the virus has been shown to promote pneumococcal transmission and disease in mouse models (Diavatopoulos et al., 2010, McCullers, 2006). The presence of viruses in the nasopharynx is thought to enhance bacterial colonisation and subsequent translocation into circulation via several possible mechanisms including: (i) disruption of the epithelial barrier (Glennie et al., 2010) (ii) up-regulation of adhesion proteins (Ganz, 2003) (iii) production of viral factors such neuraminidases which aid in exposing bacterial receptors on mucosal surfaces (McCullers and Bartmess, 2003), and (iv) impairment of the host immune system components (Kukavica-Ibrulj et al., 2009, McNamee and Harmsen, 2006). Mice infected with influenza virus showed increased susceptibility to pneumococcal infection associated with impairment of neutrophil function and enhancement of cytokine production (McNamee and Harmsen, 2006). A similar study in mice, using human metapneumovirus (hMPV) also predisposed virus-infected mice to severe pneumococcal infection, which was accompanied by increased levels of inflammatory cytokines and chemokines (Kukavica-Ibrulj et al., 2009). Excessive production of cytokines can lead to local inflammation and tissue damage, which could be a factor for increased susceptibility to secondary bacterial infections. Such excessive production of pro-inflammatory cytokines is also experienced during HIV infection and is responsible for damaging the mucosal epithelial barrier, leading to persistent bacterial colonisation and disease (Glennie et al., 2010). A recent study, in humans, highlighted the importance of taking age into account when interpreting the interaction between bacteria, and bacteria and viruses in carriage (Rodrigues et al., 2013a). The study demonstrated a significant age-independent association between viral infection in the upper respiratory tract and detection of *S. pneumoniae* and between viral infection and viral load and *S. pneumoniae* colonisation density.

However the study was not able to demonstrate any association in carriage between the three bacteria understudy: *S. pneumoniae*, *H. influenzae*, and *S. aureus* (Rodrigues et al., 2013a). The summary of microbial species interactions in the nasopharynx of a healthy carrier is shown below (Figure 1.15).

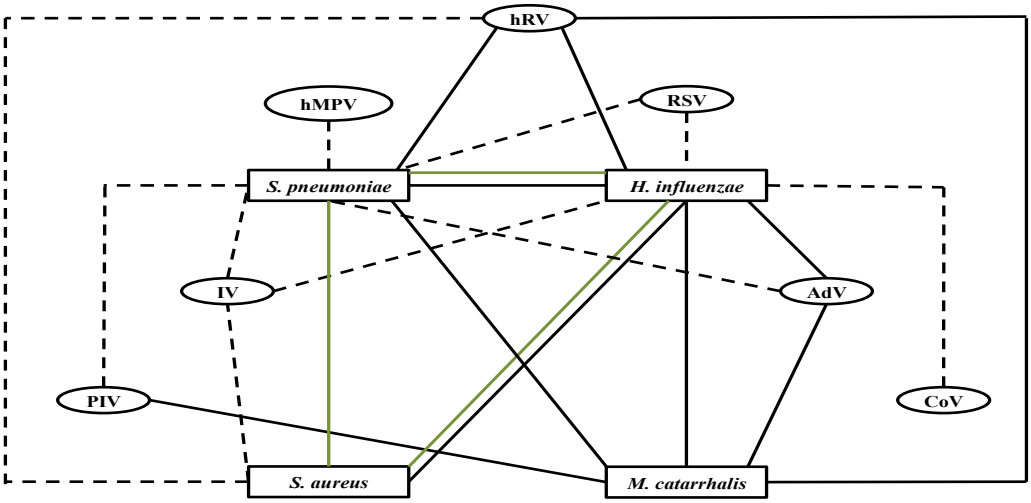


Figure 1.15 Model of bacterial and viral interaction in the nasopharynx

The proposed model is based on results from carriage experiments conducted on mouse models and *in-vitro* studies. It depicts the cumulative dynamics of bacterial and viral interactions during asymptomatic nasopharyngeal carriage as reported in all cumulative literature sources. The interactions shown involve four common nasopharyngeal bacteria: *Streptococcus pneumoniae*, *Haemophilus influenzae*, *Staphylococcus aureus*, *Moraxella catarrhalis* and seven common respiratory viruses: rhinoviruses (hRV), respiratory syncytial virus (RSV), adenoviruses (AdV), coronavirus (CoV), influenza viruses (IV), para-influenza viruses (PIV), and human metapneumovirus (hMPV). The green lines represent a negative association (competition), while black lines represent a positive association (synergism). For black dotted lines, evidence of associations is only available from *in vitro* and/or

animal studies. The model shows that viruses and bacteria were engaged in positive association only, perhaps the reason for the observed synergism during bacterial and viral co-infection. Bacterial-bacterial interactions on the other hand involved both positive and negative association (Bosch et al., 2013).

1.10. Impact of vaccination and antimicrobial drugs on microbiota

Therapeutic and vaccine interventions can alter the microbial interactions in the nasopharynx. Several studies have reported competitive associations between *S. pneumoniae* and *S. aureus* (Xu et al., 2012), raising fears that the introduction of PCV7 might increase carriage of *S. aureus* and disease post-vaccination (Madhi et al., 2007b). This has not been the case and on the contrary, PCV7 has been associated with an increase in nontypeable *H. influenzae* in acute otitis media aetiology (Dunne et al., 2013), suggesting *H. influenzae* might become a more common cause of AOM in children exposed to pneumococcal conjugate vaccination (Xu et al., 2012). In contrast, a 10-year prospective study conducted in children predicted that PCV usage could result in less otitis media by *H. influenzae* given the demonstrated association with chronicity arising froma mixture of *S. pneumoniae* and Nontypeable *H. influenzae* (Dagan et al., 2013). It is evident from the above findings that there is a shift in microbial dynamics in carriage under vaccine selection pressure, and it is anticipated that PCV13 could result in a similar change in carriage dynamics. Currently, data on the total microbial profiles in the nasopharynx and how that varies with conditions such as HIV infection are limited, and no such data exist for Malawi.

1.11. Detection of the human microbiota by 16S rRNA gene analysis

Conventionally, culture-dependent techniques have for long been employed to identify microbial species in different environments. These techniques limited the range of organisms to be detected to only those that would actively grow in a laboratory culture and made distinguishing the individual components of a microbiota challenging. Thus a large proportion of uncultured or low abundance microorganisms remained unknown. To overcome this limitation, molecular and culture-independent techniques have been employed (Olsen et al., 1986). These techniques analyse the DNA/RNA extracted directly from the samples to allow investigation of several aspects of microbial communities such as number of microbial species present in a given community (taxonomic diversity) as well as describing the biological functions which each member of a given community performs (functional metagenomics) (Morgan and Huttenhower, 2012). The commonly used target for determining microbial diversity is the 16S ribosomal RNA (16SrRNA) gene, which is universal in all bacteria (Chakravorty et al., 2007). The use of 16S rRNA gene for bacterial classification has found widespread application in clinical practice (Clarridge, 2004) as well as scientific investigations (Petti et al., 2005). The 16S rRNA gene sequence is about 1.5 kbp long and contains both the conserved and hypervariable regions (Figure 1.16) (Drancourt et al., 2000).

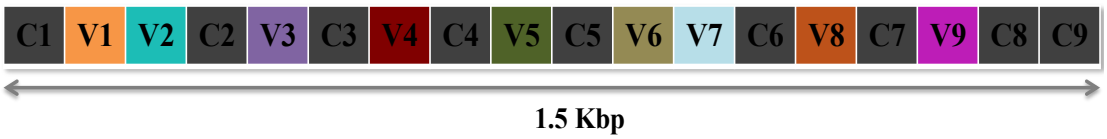


Figure 1.16 Conserved and hypervariable regions of the 16S rRNA gene

Conserved regions (C1-C9) are shown in grey while variable regions (V1-V9) appear in different colours (Petrosino et al., 2009).

There are nine hypervariable regions (V1-V9) reported and these show considerable sequence diversity among different bacterial species (Chakravorty et al., 2007). The diversity of the 16SrRNA sequences forms the basis for differentiation between organisms at the genus or indeed species and subspecies level (Drancourt et al., 2000). Although this diversity is good for identification, there is no single hypervariable region that is able to distinguish among all bacteria (Chakravorty et al., 2007).

1.11.1. Identification of microbiota by 454 sequencing of 16S rRNA gene

Sequencing of nucleic acid involves elucidation of the exact order of nucleotides present in a given DNA or RNA molecule. Large scale sequencing has traditionally been dependent on the automated sanger sequencing method developed by Fred Sanger almost 3 decades ago where sequencing was performed using chain terminating inhibitors (Sanger et al., 1977). The Sanger sequencing method, regarded as the first-generation sequencing, was central to the landmark human genome project. However the human genome project was very costly ($3 billion) and took 13 years to complete in 2003 (Grada and Weinbrecht, 2013). This triggered a shift towards sequencing technologies that would provide low cost and accurate genomic data within a day (Metzker, 2010). Thus the next generation sequencing (NGS) technologies were developed and these offer the advantage of providing a cheaper and higher-throughput alternative to sequencing than the traditional Sanger method (Metzker, 2010). Several NGS platforms have been developed and these include Roche/GS FLX Titanium, Illumina/Solexas GA_{II}, Life Technologies/Ion Torrent Personal Genome machine (PGM), Helicose Biosciences and Pacific Biosciences. Each of these platforms performs sequencing differently, however, they share some

generalized procedures such as template preparation, sequencing and imaging, and data analysis (Metzker, 2010). Template preparation involves two different methods: clonally amplified templates from single DNA molecules, and single DNA molecule templates. Clonally amplified templates are PCR dependent and the reaction may be conducted either in a liquid medium (emulsion PCR) (Dressman et al., 2003) or solid-phase (Fedurco et al., 2006). Single DNA molecule templates are immobilized on the solid support either by the template itself or by the polymerase enzyme, however PCR amplification is not required (Metzker, 2010). For the determination of microbial species using 16SrRNA genes, 454 pyro-sequencing is commonly used (Lee et al., 2013b, Quince et al., 2009). The GS-20 sequencer, with read lengths of ~100bp, was the first commercially available next generation sequencer released in 2005 (Margulies et al., 2005). With the current Genome Sequencer FLX system with long-read GS FLX titanium chemistry, read lengths of 400-600bp are possible (Rothberg and Leamon, 2008). The 454 sequencing is based on sequencing by synthesis technology of clonally amplified templates. This technology utilises an enzymatic system consisting of DNA polymerase, ATP sulfurylase, and luciferase. The incorporation of each nucleotide by DNA polymerase is accompanied by the release of inorganic pyrophosphate (PPi). The free PPi acts as a substrate for ATP sulfurylase and luciferase enzymes in a coupled reaction, which results in the production of light (Nyren, 1987) (Figure 1.17). The amount of light produced is proportional to the number of nucleotides incorporated.

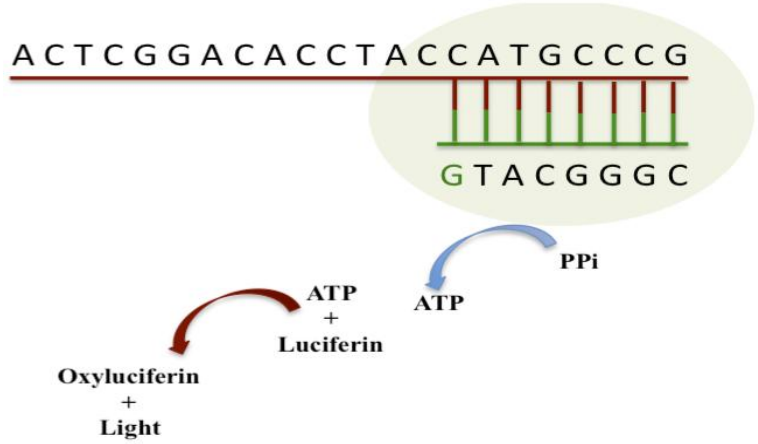


Figure 1.17 The principle of sequencing-by-synthesis

The template DNA strand to be sequenced is represented in red, the annealed primer is shown in green and the DNA polymerase is shown as the green oval. The incorporation of a complementary base (the green "G") generates inorganic pyrophosphate (PPi), which is converted to ATP by the sulfurylase (blue arrow). Luciferase (red arrow) uses the ATP to convert luciferin to oxyluciferin, which is detected as light (Rothberg and Leamon, 2008).

The 454 sequencing machinery consists of an assembly containing (a) deoxynucleotide triphosphates (dNTPs) and (b) a flow cell that includes a fiber-optic slide containing wells. The wells are contained on the PicoTiterPlate and are loaded with capture beads containing amplified PCR strands plus numerous small beads with enzymes. Fitted with a liquid system, the 454 machinery is able to sequentially wash the PicoTiterPlate with dNTPs and other necessary reagents. Finally (c), an imaging assembly fitted with CCD camera to image the fiber-optic slide, and allows the capture of emitted light from the bottom of each individual well, and a computer for instrument control operations and visualisation (Figure 1.18) (Rothberg and Leamon, 2008).

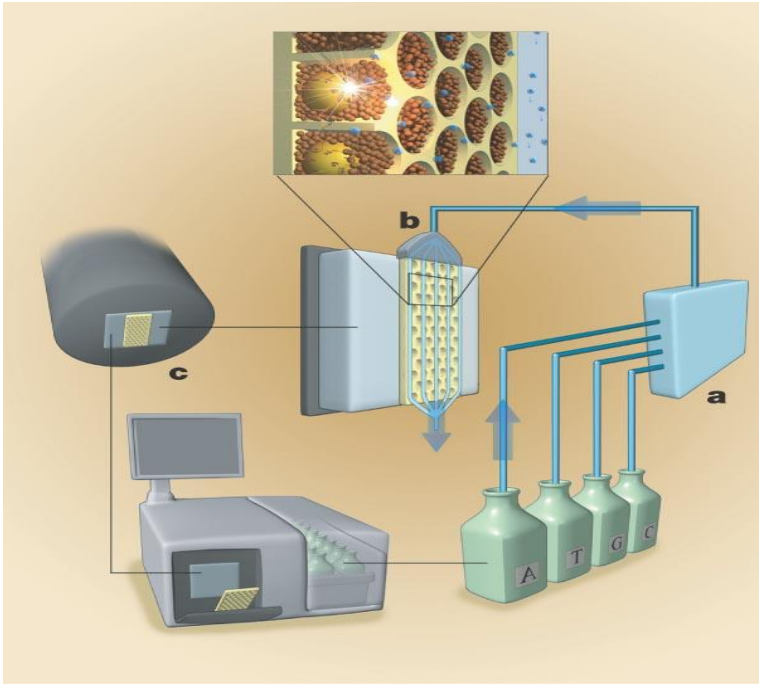


Figure 1.18 The 454 sequencing machine assembly

The assembly consists of (a) a set of bottles containing the four dNTPs, (b) a flow cell that includes a fiber-optic slide containing wells and (c), an imaging assembly fitted with CCD camera to image the fiber- optic slide, and a computer for instrument control operations and visualisation (Figure 1.18) (Rothberg and Leamon, 2008, Margulies et al., 2005).

The 454 sequencing occurs in three major steps namely DNA library preparation, Emulsion PCR amplification and Sequencing. During library preparation, the isolated genomic DNA is fragmented (Figure 1.19A), ligated to adapters, which are color coded green and red (Figure 1.19B) and then separated into single strands (Figure 1.19C). The adapters are essential for purification, amplification and sequencing.

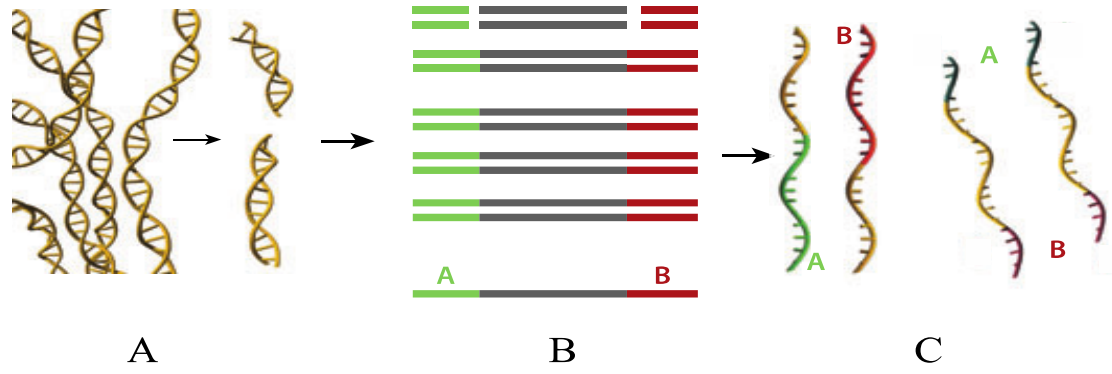


Figure 1.19 DNA library preparation

The fragments containing both A and B (A/B) adapters are selected using avidin-biotin purification and used in the PCR amplification step (Rothberg and Leamon, 2008).

The single stranded DNA fragments are bound to their own unique beads and then compartmentalized in the droplets of a PCR-reaction-mixture-in-oil emulsion (Figure 1.20, A and B). PCR amplification occurs within each droplet, resulting in each bead carrying multiple copies of a unique DNA template (Figure 1.20 C). Once amplification is complete, the oil emulsion is broken and the DNA strands on the beads are denatured (Figure 1.20 D).

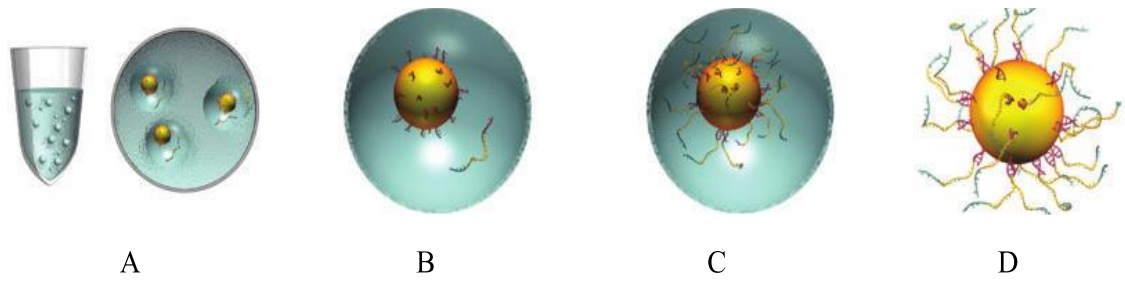


Figure 1.20 Emulsion PCR amplification of single stranded DNA

(Rothberg and Leamon, 2008).

Each bead carrying clonally PCR-amplified DNA fragments is deposited into a unique well on the fiber-optic-slide together with smaller beads carrying immobilized enzymes required for a solid phase pyrophosphate sequencing reaction (Figure 1.21)

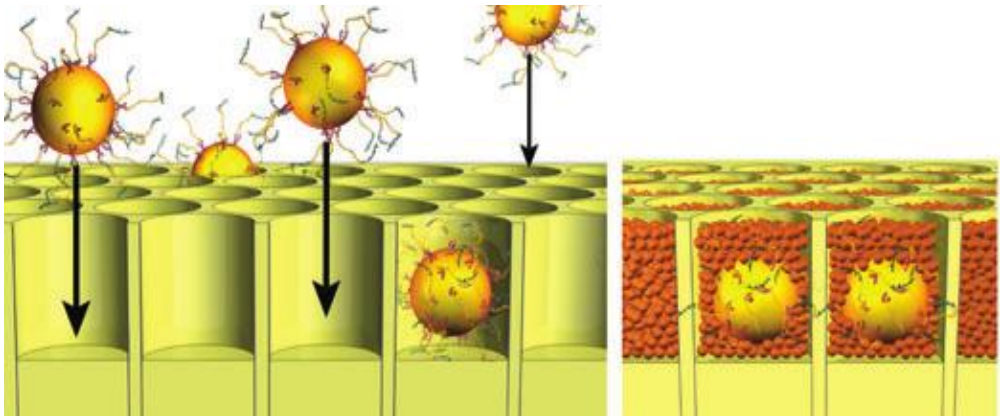


Figure 1.21 454 Sequencing of amplified single stranded immobilised DNA

A single bead containing clonally amplified DNA is deposited per well. On average, each well has a diameter of 44um. About 400, 000 reads are obtained in parallel (Rothberg and Leamon, 2008).

1.11.2. Bioinformatic tools for analysing 16S sequences

Pyrosequencing of PCR amplicons has led to an in-depth understanding of environmental microbial diversity (Bates et al., 2011, Lee et al., 2013b) and the human microbiome (Bogaert et al., 2011, Consortium., 2012). This is due to high read depth where multiples of individual sequence reads are generated simultaneously, which provides a tool for estimating the composition and relative abundance of microbial operational taxonomic units (OTUs) in a given community. However, due to the large number of sequence reads generated and the lack of

consensus reference sequences, distinguishing false from true sequence diversity has proven difficult. Thus data generated through 16S amplicon pyrosequencing has the potential for overestimating species abundances due to artificial sequences arising from PCR single base substitutions, PCR chimeras and sequencing errors (Quince et al., 2011). Chimeras are hybrid products between multiple parent sequences that are generated when incomplete extension occurs during the PCR process and the resulting fragment effectively acts as the primer in the next round of PCR (Figure 1.22) (Quince et al., 2011).

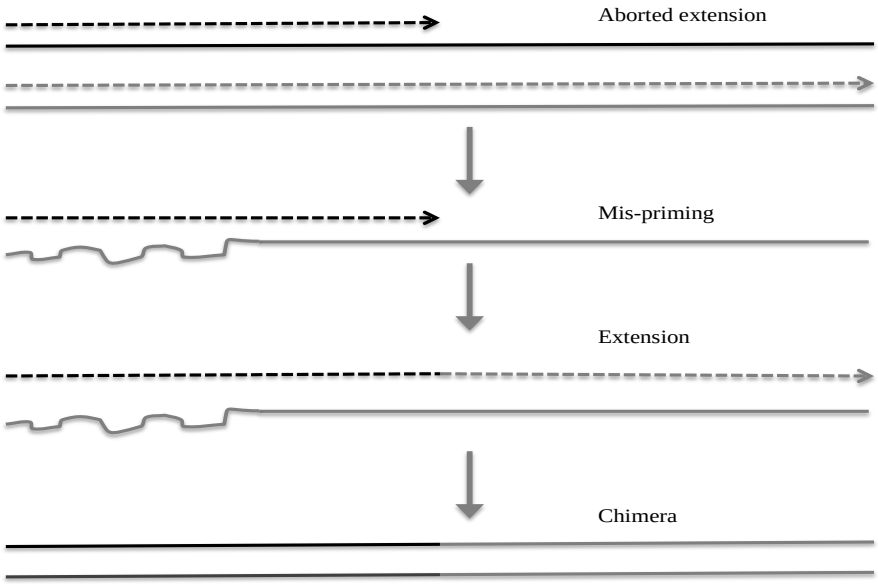


Figure 1.22 PCR generated chimeric sequences

Formation of chimeric sequences during PCR amplification of 16S rRNA gene. An aborted extension product from an earlier cycle of PCR can function as a primer in a subsequent PCR cycle. If this aborted extension product anneals to and primes DNA synthesis from an improper template, a chimeric molecule is formed (Haas et al., 2011).

To minimise the impact of sequencing errors on the interpretation of microbial diversity, several computational approaches have been developed: (i) removing sequence reads with sequencing errors such as ambiguous base calls and mismatches to primer sequences (Huse et al., 2010, Quince et al., 2011); (ii) trimming regions of the sequence with low quality scores (Chou and Holmes, 2001); (iii) *flowgram-based clustering*, where information obtained by clustering pyrosequencing flowgrams is incorporated into the pipeline for removing error reads (Quince et al., 2011, Quince et al., 2009, Reeder and Knight, 2010). These approaches however have had very limited application because of their requirement for very high computational capabilities (Schloss et al., 2011). To minimise the impact of PCR-induced sequence artefacts on microbial diversity, several approaches have been developed to identify and remove chimeras from the analysis. Three common approaches are known. The first approach, ChimeraSlayer (Haas et al., 2011), involves searching the ends of a query sequence (30% of the length from each end) against a database of reference chimera-free 16S sequence to identify potential parents of a chimera. The second approach is Perseus (Quince et al., 2011) which does not use a reference database, however it requires a training set of sequences similar to the sequences being characterised. Perseus compares query sequences with all pairs of sequences having higher abundance is made based on the assumption that a chimera undergoes fewer rounds of amplification than its parents. The closest pair is thus selected, and its three-way alignment with the query sequence is made. Perseus is performed after AmpliconNoise, an algorithm that accounts for sequencing errors and PCR single base substitutions. The third approach, Uchime (Edgar et al., 2011) combines principles of both ChimeraSlayer and Perseus by either using a database of chimera-free sequences or detecting chimeras *de novo* by using sequence abundance data

respectively. Recently, an open-source Mothur software, a 16SrRNA data analysis tool was developed (Schloss et al., 2009). Mothur is a single software command-line computer program that is built upon other pre-existing tools for analysing sequence data from microbial communities. Basically, Mothur is used to trim, screen, and align sequences; distance matrix-based sequence clustering, assign representative sequence reads to operational taxonomic units (OTUs); and to provide an estimation of within- (α-diversity) and between- (β-diversity) sample diversity (Schloss et al., 2009).

1.12. Aims of the study

1.12.1. Main objectives

The primary objective of this study was to investigate nasopharyngeal carriage of *Streptococcus pneumoniae* in Malawian children and adults, prior to the introduction of the 13-valent pneumococcal conjugate vaccine (PCV13). *S. pneumoniae* interacts with a wide range of other microorganisms colonising the nasopharynx, which impacts on pneumococcal virulence and carriage dynamics. Therefore, this study also characterised carriage of nasopharyngeal microbiota in Malawian children and adults from matching samples. Recent microarray serotyping and 454 pyrosequencing technologies were employed to identify *S. pneumoniae* and other microbiota in carriage respectively.

1.12.2. Specific objective

This study focused on three specific objectives:

1. To characterise the general profile of pneumococcal serotypes in carriage in Malawian children and adults.

2. To assess the prevalence of simultaneous carriage of multiple serotypes (multiple carriage) in Malawian children and adults.

3. To characterise carriage and composition of nasopharyngeal microbiota in Malawian children and adults.

1.13. Significance of the study

Carriage of *S. pneumoniae* is considered the first step in the development of invasive pneumococcal disease and is also responsible for the spread of the pathogen in a given population. Several carriage studies that have been conducted have provided information that has helped in formulating disease control strategies. Most importantly, carriage studies have been used to estimate the impact of vaccination before and after vaccine introduction. Recent reports suggest carriage data can be used to (i) aid in licensing and implementation of new vaccines (Goldblatt et al., 2013) and (ii) accurately estimate post-PCV changes in IPD incidence (Weinberger et al., 2013), particularly in settings with poor surveillance system for evaluating impact of vaccination. Therefore, the importance of undertaking carriage studies cannot be overemphasised. In Malawi, pneumococcal carriage studies are inadequate and until now, there has been no report on the prevalence of multiple carriage in this

setting. In addition, this is the first time in Malawi that carriage of nasopharyngeal microbiota has been characterised. Findings from this study will fill this gap in knowledge in the Malawian setting, and provide baseline information that will aid in (i) estimating vaccine efficacy and (ii) predicting potential vaccine-induced changes in carriage dynamics post-vaccination, which is important in the assessment of the effectiveness of clinical interventions for Malawi.

1.14. Structure of the thesis

Chapter 2 of this thesis outlines methodologies used in this study ranging from sample collection through culturing to characterisation of *S. pneumoniae* by microarray and nasopharyngeal microbiome by 454 pyrosequencing. Chapter 3 outlines the result for the general pneumococcal carriage profile and the prevalence of multiple carriage in Malawian children and adults. The dynamics of pneumococcal carriage and genetic diversity overtime is presented in Chapter 4. Chapter 5 characterises capsular polysaccharide locus variants of *S. pneumoniae* detected in carriage in Malawi and provides preliminary findings on whether the variants have an altered ability to colonise and cause death in mouse models. Chapter 6 describes the nasopharyngeal microbiota in Malawian children and adults. The last chapter (Chapter 7) provides a general discussion of the findings from this study, outlines major limitations and also provides suggestions for future work to be performed in order to address some of the questions arising from this study.

Chapter 2: Materials and methods

2.0 Background

This study was conducted in Malawi (Figure 2.1), a resource poor setting with a high burden of HIV infection (Lau and Muula, 2004) and invasive pneumococcal disease (IPD**)** (Gordon et al., 2000). The study was part of an on-going project at the Malawi-Liverpool-Wellcome Trust (MLW) Clinical Research Programme (http://www.mlw.medcol.mw) in Blantyre, Southern Malawi, investigating the pathogenesis, transmission and prevention of bacterial and viral infections in Malawi.

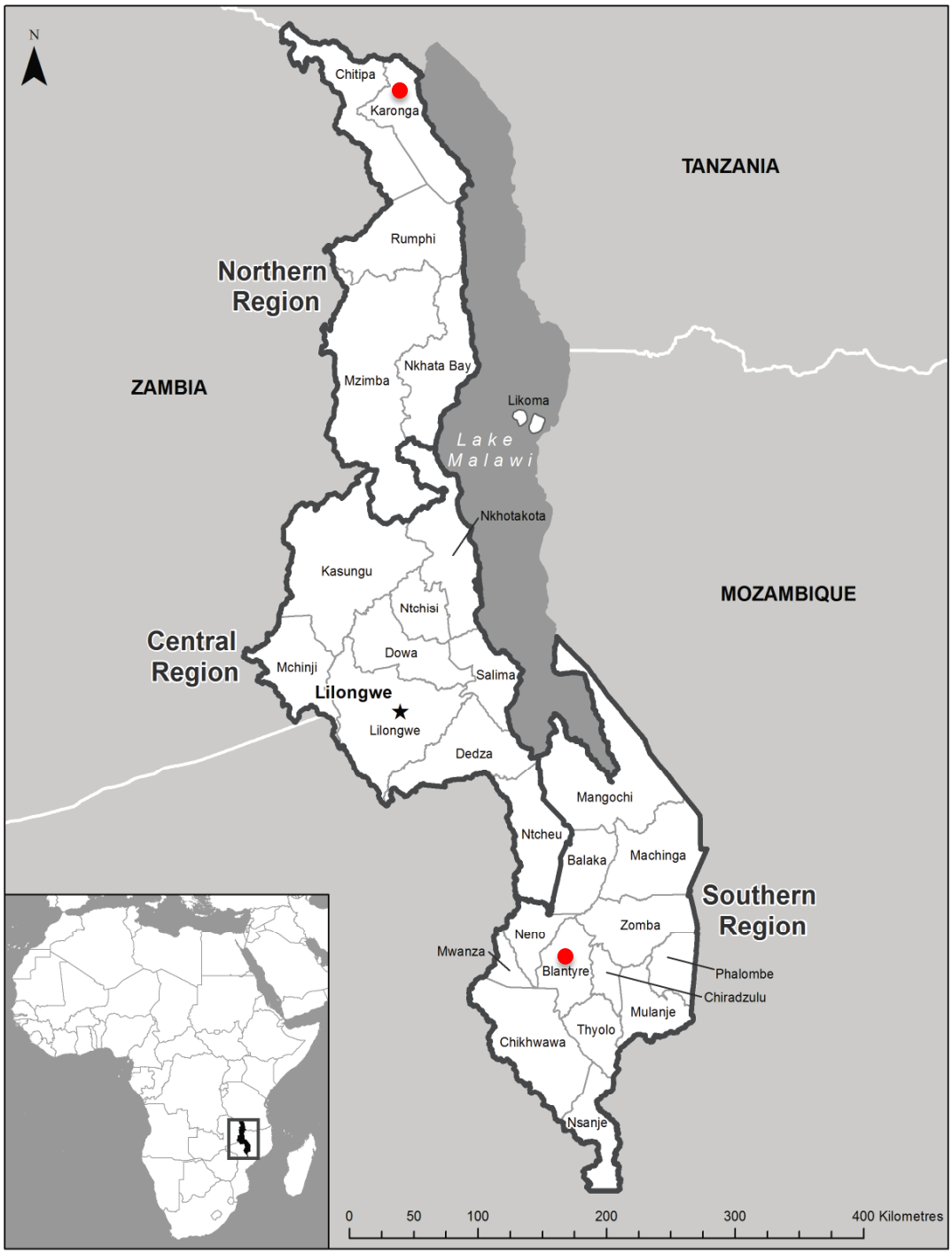


Figure 2.1 Map of Malawi

Map of Malawi showing sampling sites in Karonga (KPS) in the Northern region and Blantyre (MLW) in the Southern region (http://www.nsomalawi.mw) located 411 miles apart (http://www.worldcities.us). The red circles on the map identify the sampling sites.

The MLW offers routine diagnostic testing services to support the Queen Elizabeth Central Hospital (QECH) and has an archive of over 5000 pneumococcal strains isolated from carriage and invasive pneumococcal disease since 1996. QECH is a government funded central hospital with 1250 bed space, provides medical care to a population of over 1 million people and has an annual admission rate of 50,000 people (Everett et al., 2011). The study was also linked to Karonga Prevention Study (KPS), a controlled randomised study (CRS) investigating asymptomatic carriage and transmission of *S. pneumoniae* in households in Karonga, Northern Malawi. The underlying laboratory work was conducted in four research centers: MLW; the Wellcome Trust Sanger Institute (WTSI) (http://www.sanger.ac.uk), Hinxton, Cambridgeshire, UK; the Institute of Infection and Global Health (IGH) (http://www.liv.ac.uk/infection-and-global-health), University of Liverpool, Liverpool, UK and the Bacterial Microarray Group at St Georges Hospital (BμG@S) (http://www.bugs.sgul.ac.uk), London, United Kingdom (UK) (Figure 2.2).

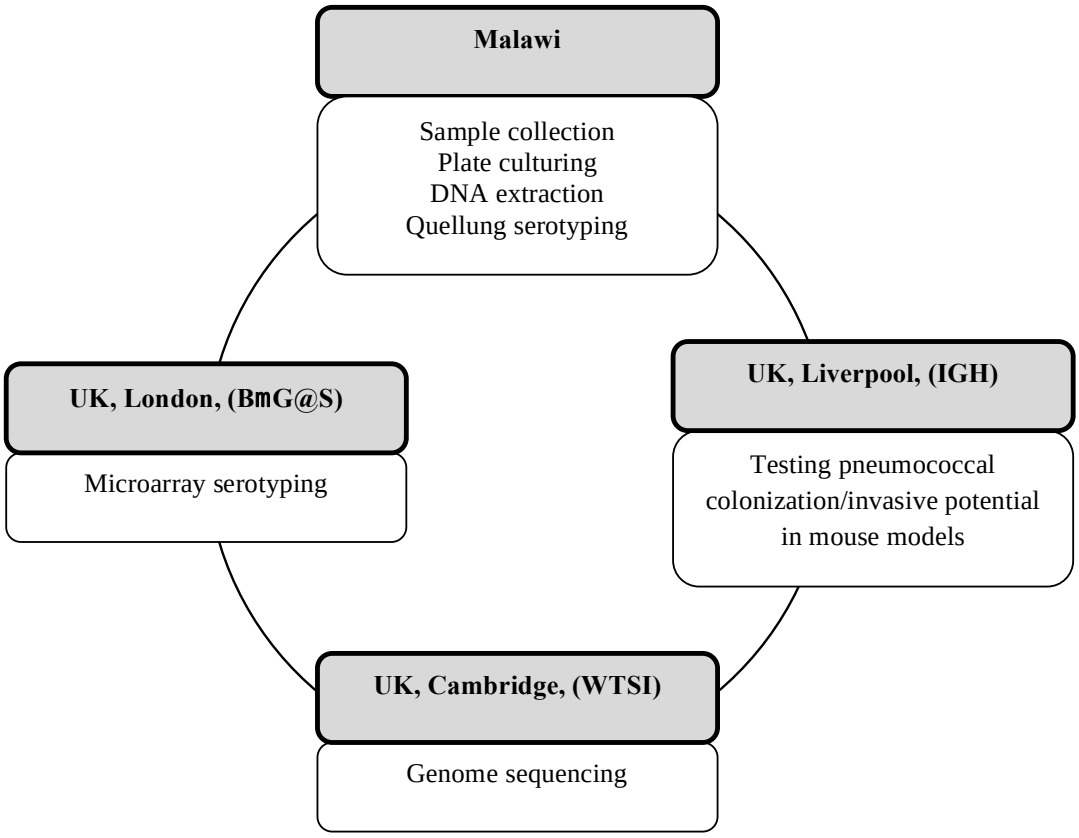


Figure 2.2 Institutions involved in the study

2.1. Sample selection in children

Samples from children were conveniently selected from a dataset collected over a period of 5 years from 2008. Only samples with a pneumococcal culture positive result were picked for this study because the aim was characterise the prevalence of multiple carriage among carriers of *S. pneumoniae*. These samples were accessed from the archives in Karonga and Blantyre (Figure 2.3). The Karonga and Blantyre sites are 411 miles apart, and this distance provided the opportunity to better understand the diversity of pneumococcal carriage in different geographical locations within Malawi, although this may be confounded by differences in age as well as HIV status. The two variables of key importance in the selection of samples were carriage of *S. pneumoniae* and HIV status of the child. One hundred and sixteen *S.*

pneumoniae culture positive samples were selected (Figure 2.3). Based on resource and time constraints, this sample size was adequate for characterising carriage of pneumococcal strains and other microbiota in Malawian children. The HIV negative children represented 62% (72/116) of the total number of children in the dataset, while 38% (44/116) was covered by HIV positive children. More HIV positive children were sampled in order to increase the number of pneumococcal carriers to address the question of how HIV infection affected carriage. This resulted in a much higher HIV prevalence compared to the reported average 10.6% for Malawi (UNAIDS, 2012a). Children under the age of five years represented 84% (97/116) of the total number of children in the dataset, and 16% (19/116) were 5 years and older. Sampling was not based on gender; however the gender distribution was similar with males representing 58% (68/116) and females 42% (48/116) of the entire children's dataset. None of the children had received the 13-valent pneumococcal conjugate vaccination now included into Malawi's expanded program on immunisation (EPI).

Karonga. Nasopharyngeal swabs were collected from a subset of infants (n=16) containing a total of 158 participating infants. Samples were collected at six weeks of age and then every four weeks for 12 months. This study analysed samples from any four subsequent sampling points, previously shown to be positive for *S. pneumoniae* in each infant. Fourteen infants had their HIV status confirmed while the status of the remaining two infants could not be established. For analyses investigating the impact of HIV infection, the two samples with unknown HIV status were not included. These two samples were only included in the analysis of pneumococcal carriage dynamics (Chapter 4), where HIV status was not of key importance. Initial processing of samples to identify pneumococcal growth was done in Karonga using

conventional techniques of culture on sheep blood agar and gentamicin (SBG) and subsequent demonstration of optochin sensitivity amongst alpha haemolytic streptococcal species. Following selection of samples, nasopharyngeal swab tips in STGG, were then shipped (at -20 °C) to Blantyre for further evaluation. Each of the infants contributed samples from four consecutive sampling points, confirmed to contain *S. pneumoniae* by Quellung reaction (Section 1.7.1), giving a total of 64 nasopharyngeal swabs. The longitudinal samples were analysed to provide an insight into and to improve understanding of carriage dynamics in children over time (Chapter 4).

Blantyre. Samples (n=102) were collected from children reporting to QECH as outpatients. These were recruited in two different clinical studies conducted at MLW. The first study (H1N1) conducted an in-depth hospital-based on-going influenza surveillance study since late 2010, designed to determine the severity and hospital case fatality due to seasonal influenza and H1N1 influenza virus. The H1N1 study recruited both adults and children presenting to the QECH with influenza-like illnesses (ILI's) and severe acute respiratory infections (SARI's). For the purposes of this study, only children under the age of 5 years presenting with ILI's and not SARI's were included. SARI's were excluded to avoid carriage estimation bias because of the known synergism between influenza infection and pneumococcal colonisation (Diavatopoulos et al., 2010). All the ILIs sampled were negative for influenza. From a total of 154 children with ILI's, 49% (76/154) were under the age of five years and these were then screened for the presence of *S. pneumoniae*. A positive pneumococcal culture was detected in 72% (55/76) of the total samples screened. Two samples did not grow on subsequent culturing and therefore only 53

were available for further analysis. A further 49 samples were drawn from a second study, which investigated the B Cellular Immunological (BCIM) responses to pneumococcal infections by HIV status and usage of antiretroviral therapy (ART). BCIM samples comprised of both longitudinal (n=16) and cross-sectional (n=33) sets. Longitudinal BCIM samples were collected from asymptomatic HIV positive children, at two sampling points: pre-ART and 12 months post-ART. BCIM samples provided an opportunity for understanding changes in pneumococcal carriage dynamics amongst HIV infected subjects following ART mediated immune reconstitution.

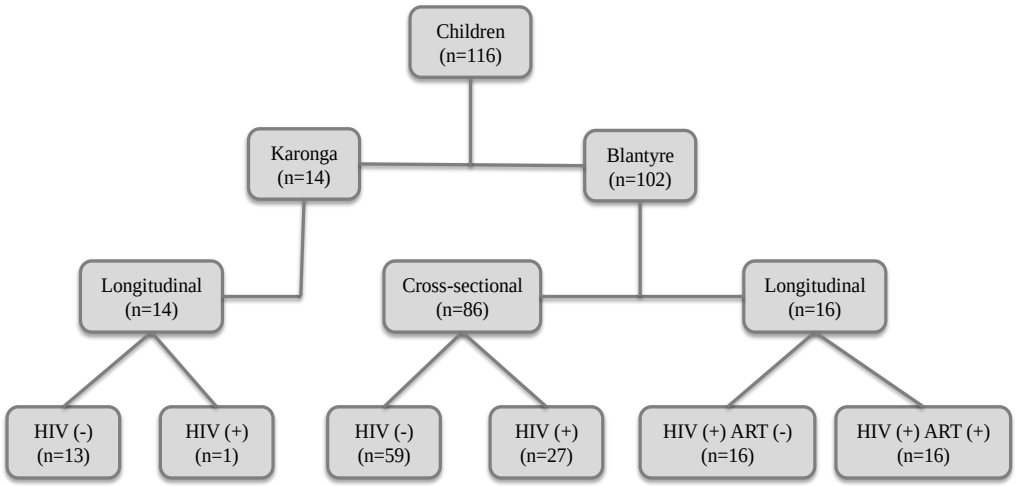


Figure 2.3 Sample selection in children

All samples were collected from children (n=116) using nasopharyngeal swabs. Four swabs were taken from each participant in a longitudinal study. The total number of nasopharyngeal swabs analysed was 200.

2.2. Sample selection in adults

Adult samples (Figure 2.4) were collected from a larger cross-sectional study in Blantyre, investigating the impact of 7-valent pneumococcal conjugate vaccine (PCV 7) and ART on invasive pneumococcal disease in Malawian adults. Samples were collected from adults reporting to QECH as outpatients from 2006 to 2008. All samples (n=117) that were culture positive for *S. pneumoniae* were selected. These samples were grouped into HIV negative (n=26) and HIV positive (n=91). The high number of HIV positive subjects could possibly highlight the higher rates of pneumococcal carriage in HIV positive compared to HIV negative individuals (Jordano et al., 2004).

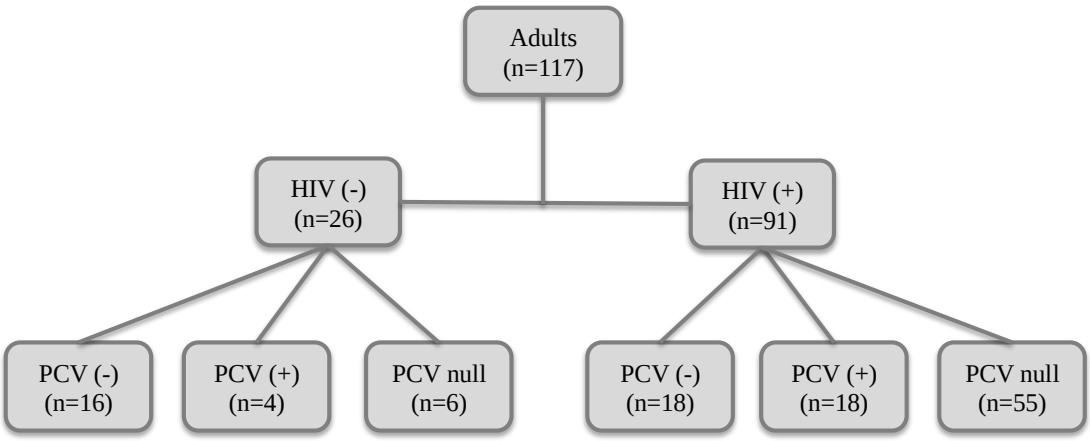


Figure 2.4 Sample selection in adults

All samples were collected from adults (n=117) recruited in a cross-sectional data set using nasopharyngeal swabs. The subjects on vaccine were subjected to the 7-valent pneumococcal conjugate vaccine (PCV7). Subjects with no available information on PCV7 usage were designated PCV null.

2.3. Collection of nasopharyngeal samples

Nasopharyngeal (NP) samples were collected by trained nurses following the WHO protocol for the detection of the pneumococcus in carriage (O'Brien and Nohynek, 2003). Briefly, the subjects' head was gently tilted back with the chin kept steady (Figure 2.5). A sterile Dacron-tipped nasopharyngeal swab (Medical Wire and Equipment, Corsham, United Kingdom) was inserted into the nasopharynx, rotated at 180 °C and then slowly removed after saturating the tip by leaving it in place for 5s. Using sterilised scissors, the NP swab tips were aseptically cut off into a labelled 2.0ml vial containing 1.0ml of skim milk, tryptone, glucose, and glycerine (STGG) medium (Gibson and Khoury, 1986), and then stored at -80 °C until required. The recovery of pneumococcal strains stored under these conditions is comparable to the recovery from direct plating (O'Brien et al., 2001).

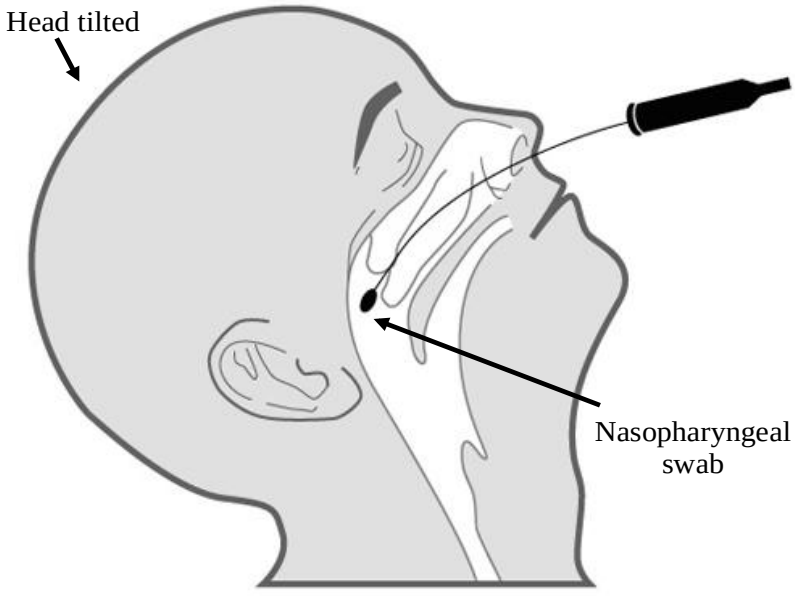


Figure 2.5 Nasopharyngeal specimen collection

(http://www.stanfordlab.com)

2.4. Culturing of *S. pneumoniae*

Nasopharyngeal swabs kept at -80 °C and stored in STGG were briefly thawed to ~4 °C and immediately pre-screened for the presence of *S. pneumoniae* by culturing an aliquot (20µl) on SBG. The SBG was prepared following the MLW Standard Operating Procedures (Appendix 1(a) and 1(b)). Positive identification of *S. pneumoniae* was by colony morphology and sensitivity to optochin (Oxoid Limited, Basingstoke, Hampshire, UK) by disk diffusion (Figure 2.6). Optochin disks are made of a water insoluble and sterilized filter paper base (8mm in diameter), impregnated with a water-soluble ethylhydrocupreine hydrochloride as an active chemical component (Bowers and Jeffries, 1955). Optochin sensitivity differentiates *S. pneumoniae* from other α-hemolytic Streptococcus species such as *Streptococcus viridans*, which are optochin-resistant (Bowers and Jeffries, 1955). A confirmatory bile (sodium deoxylate) solubility test is often performed to distinguish *S. pneumoniae* from other alpha-hemolytic Streptococcus species that are sensitive to optochin (Arbique et al., 2004a). STGG samples containing *S. pneumoniae* were re-cultured in a series dilution for subsequent analysis. A series dilution was performed to obtain a culture with non-overlapping growth of colonies, to ensure a less competitive environment for growth of less abundant serotypes. Briefly, cryotubes containing samples in STGG media were vortexed for 10-20 seconds. A series dilution (1, 1/10, and 1/100) of STGG suspension was prepared in brain heart infusion (BHI) broth (Oxoid, UK). The suspension (50µl) was pipetted onto a selective colistin oxolinic acid blood agar (Oxoid, UK) plate and spread evenly. The plates were incubated overnight at 37^{o}C supplemented with 5% carbon dioxide. An enhanced carbon dioxide environment is critical for the isolation of *S. pneumoniae*, because it leads to the production of large mucoid pneumococcal colonies, which can

easily be identified (Howden, 1976, Austrian and Collins, 1966). The dilution with adequate non-overlapping growth of colonies was taken for further processing.

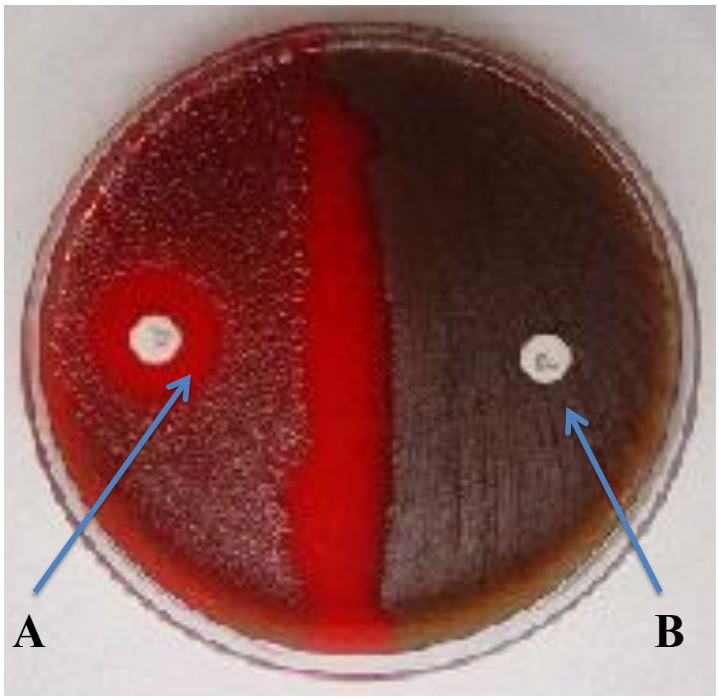


Figure 2.6 Plate culturing of *S. pneumoniae*

S. pneumoniae is cultured on selective gentamicin blood agar or COBA media. An inhibition zone of ≥15mm (Gardam and Miller, 1998) around an optochin disk (A) is a positive identification of *S. pneumoniae*, while resistance to optochin, shown by lack of an inhibition zone (B) signifies presence of an organism other than *S. pneumoniae*.

2.5. DNA extraction

Culture dependent DNA extraction for microarray serotyping. Sterile phosphate buffered saline (PBS) (Sigma-Aldrich, Germany) (1ml) was added to the culture plate. Using a spreader, a suspension of all colonies was made by scraping the whole plate. The resulting suspension was transferred to a 1.5ml micro-centrifuge tube (Eppendorf, Hamburg, Germany) using a 1ml pipette (Starlab, UK). The tubes were subjected to centrifugation at 7500rpm for 10minutes to pellet the bacteria. DNA was extracted from the bacteria pellet using QIAamp DNA mini kit (Qiagen, Germany) based on the manufacturer's protocol as follows: the bacterial pellet was re-suspended in lysozyme 180µl solution (20mg/ml, in buffer containing 20mM Tris-HCl, pH 8, 2mM EDTA, 1.2% Triton) and incubated for 60minutes at 37 °C. EDTA (ethylenediamine tetra-acetic acid) is a chelating agent that binds to metallic cofactor ions required by DNA degrading enzymes. This inhibits the enzymatic activity of these enzymes ensuring high DNA yields. Triton is an anionic detergent that denatures cell membrane proteins and disrupts bacterial cells for easy DNA extraction. Proteinase K (20µl) and Buffer AL (200µl) were added to the suspension, vortexed and then incubated at 56 °C for 60minutes. RNase A (100mg/ml, 4µl) was added and incubated for 2 minutes at room temperature, followed by a further addition of Buffer AL (200µl) and incubation at 70 °C for 10 minutes. Proteinase K and RNase A are necessary for digesting proteins and RNA, which usually contaminate DNA preparations. Absolute ethanol (200µl) was added to the solution and mixed by manually turning the micro-centrifuge up and down a few times, to ensure adequate precipitation of the DNA. The ethanol solution was added to the spin column followed by centrifugation at 8000rpm for 1 minute. The column was washed with buffer AW1 (500ul) and centrifuged at 8000rpm for 1 minute followed

by a second wash with buffer AW2 (500µl) and centrifuged at 13,000rpm for 3minutes.The flow through was discarded and the column was placed in a new tube after each wash. Residual ethanol was removed by performing a further centrifugation at 13,000 rpm for 1 minute. The column was placed in a clean 1.5ml tube. Molecular biology grade water (200µl) was added onto the column membrane and allowed to stand for 1 minute at room temperature, to ensure the DNA was fully solubilised. The DNA was eluted from the column by centrifugation at 13,000 rpm for 1 minute. The DNA was quantitated by NanoDrop (ND 1000, Labtech, UK) and then stored at 4 °C.

Culture dependent DNA extraction for whole genome sequencing. Colonies with different morphologies were picked from the plate. Each colony was inoculated into brain heart infusion (BHI) broth (9ml) and then incubated overnight at 37 °C in 5% carbon dioxide. The cells were harvested by centrifugation at 5000g for 5.0 minutes. Genomic DNA was extracted using QIAamp DNA mini kit as before. To determine the pneumococcal serotypes, the DNA was subjected to PCR amplification using CDC multiplex PCR primers (Pai et al., 2006) (Appendix 7 (a)) as follows: 95 °C for 15 mins, followed by 35 cycles of 94 °C for 30 secs, 54 °C for 90 secs, 72 °C for 60 secs, followed by 72 °C for 10 min and then a 4 °C hold. The PCR amplicons were analysed on a Bioanalyser (2100, Agilent Technologies, USA). The PCR step was only conducted for pneumococcal serotypes isolated from samples with multiple carriage. The DNA for PCR confirmed serotypes were stored at 4 °C until further analysis by whole genome sequencing.

2.6. Molecular serotyping of *S. pneumoniae* by microarray

This study used the *Streptococcus pneumoniae* capsular polysaccharide (SP-CPS) microarray assay (Chapter 1, section 1.7.2.2.1), which was developed by the bacterial microarray group at St George's (BμG@S) in London, to detect the serotypes in carriage. The BμG@S SP-CPS microarray also offers enhanced utility for novel serotype discovery; nontypeable (NT) strains investigation, detection of multiple serotype carriage and performing comparative genomic hybridisation (arrayCGH) using spTIGR4+ R6 genome backbone (Hinds et al., 2009). The limit of detection for the BμG@S SP-CPS microarray is 1% in samples carrying multiple serotypes (Hinds et al., 2009).

2.6.1. DNA Labelling

The Kreatech Universal Linkage System (Kreatech ULS) is a non-enzymatic labelling technique consisting of a square planar platinum complex stabilised by a chelating diamine, a detectable molecule (marker) such as enzyme or fluorochrome, and a leaving group (X) such as chloride or nitrate ion (Figure 2.7A). The leaving group is displaced during binding to the N7 position of guanine moieties in nucleic acids, to form a co-ordinate bond (Figure 2.7B) (van Gijlswijk et al., 2001, Wiegant et al., 1999).

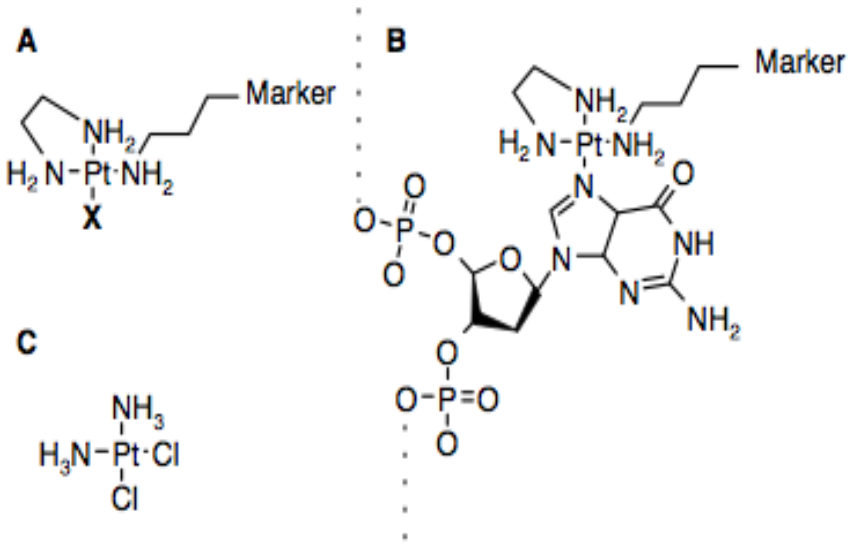


Figure 2.7 ULS square planar platinum complex

The complex is stabilised by a chelating diamine (A) and it binds to the N7 position of guanine moieties in nucleic acids, to form a co-ordinate bond (B). (van Gijlswijk et al., 2001).

In this study, the purified DNA was heat fragmented by incubating at 95 °C in a thermocycler (Eppendorf, Germany) for 30 minutes, with a final temperature cycle of 4 °C for 3 minutes. The DNA was labelled with ULS containing cyanine green (ULS-Cy3) and red (ULS-Cy5) fluorescent, maintaining a ratio of 1µl ULS dye per 1ug DNA in every ULS-Cy3 or ULS-Cy5 labelling. The DNA was incubated at 85 °C for 30 minutes in a thermocycler with a final holding step at 4 °C for 3 minutes. The labelled DNA was purified using KREApure column to remove non-reacted ULS-Cy, which increase background noise if not removed before the hybridisation step.

2.6.2. Sample preparation, loading and hybridisation

The purified Cy3/Cy5 labelled DNA (18 μl) was mixed with Agilent blocking agent (10x, 4.5 μl) and Agilent hybridisation buffer (2x, 22.5 μl) in a 0.5ml tube. The mixture was incubated at 95 °C for 5 minutes followed by 37 °C for 30 minutes. The hybridisation assembly (http://www.agilent.com) was prepared based on manufacturers' protocol as follows: A clean gasket slide (8x15K) was placed in a hybridisation chamber base with the gasket label facing up, and properly aligned with the rectangular section of the chamber base. The hybridisation solution (40μl) was slowly dispensed onto each of the eight gasket wells in a "drag and dispense" manner (Figure 2.8).

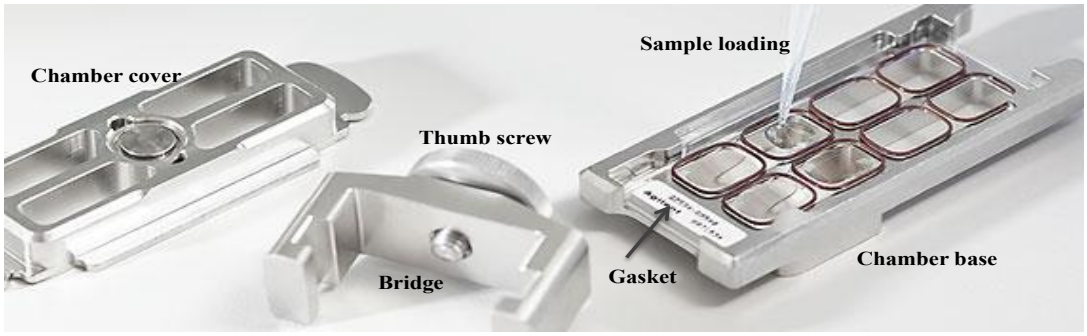


Figure 2.8 Microarray hybridisation assembly

The microarray slide was placed array side down onto the gasket slide, to form an array-gasket sandwich. A chamber cover was placed on top of the array slide, and gently clamped with the bridge to firmly hold the assembly together. The assembled chamber was vertically rotated to assess the mobility of the bubbles. Stationary bubbles were moved by gently tapping the assembly on a hard surface. The hybridisation chamber (Figure 2.9) was placed in a rotator rack in the hybridisation oven (set to 65 °C) and rotated overnight at 20rpm.

Figure 2.9 Hybridisation chamber

2.6.3. Slide washing and scanning

The array-gasket sandwich was removed from the hybrid chamber base and quickly transferred to a trough containing Oligo aCGH buffer 1 at room temperature. The array–gasket sandwich was opened whilst submerged in wash buffer 1.The array slide was then transferred to the rack in second trough of wash buffer 1 and stirred for 5 minutes. The final wash was performed by transferring the array slide rack to trough containing wash buffer 2 at 37 °C and stirred for 1 minute. The slide rack was gently removed from wash buffer 2 to minimise droplets on the slides. The slide with Agilent barcode facing up was placed in a slide holder and then scanned immediately. The developed microarray slide was scanned using High-Resolution Microarray Scanner (Agilent Technologies, USA). The raw microarray image files were automatically read and processed using Agilent's feature extraction software (version 10.7) (http://www.genomics.agilent.com).

2.6.4. Serotype identification

The output from microarray scanner was analysed by a previously described empirical Bayesian model (Newton et al., 2011), which determines serotypes present in a given sample by calculating the probabilities of different serotype combinations. The relative abundance of each serotype in a sample with multiple serotypes was determined using a Bayesian random effects model (Newton et al., 2011) based on different fluorescence intensities. The microarray serotype data was analysed using Microsoft excel and Stata®11.0 (appendices 2(a) and 2(b)). Following is an illustration of the output from microarray, for single and multiple carriage episodes involving serotypes 3 and 14.

2.6.4.1. Single carriage identification of the serotype 3

To establish the pathogen present in a sample, microarray utilises PathID probes. In sample 11787, the fluorescence intensity for the pneumococcus was over 10000 fluorescence units (Figures 2.11) while all the other pathogens were observed at the zero-baseline. This suggested that only *S. pneumoniae* was present in all colonies obtained from the culture plate. The pneumococcal serotype in the sample was determined using STID probes, which detect the serotype specific genes present at the CPS locus. The CPS gene locus products are classified into 249 homology groups (HGs) based on amino acid sequence similarity and predicted function (Bentley et al., 2006, Aanensen et al., 2007). The CPS locus for wild type pneumococcal serotype 3 consists of eight genes and is about 10, 337bp long (Figure 2.10) (Bentley et al., 2006). In sample 17787, detection of homology group presence (HGID) showed eight HGs (Figure 2.12): 0.01 [*wzg*], 1.02 [*wzh*], 2.02 [*wzd*], 3.02 [*wze*], 20.07 [*ugd*], 196.01 [*wchE*], 197.01 [*galU*], and 198.01 [*pgm*], representing eight

CPS genes for serotype 3 (Figure 2.10). The presence of serotype3 in the sample was further confirmed by STID analysis. The intensities of all possible serotypes were at the zero-baseline, however, the elevated signal intensities for serotype 3 (>30000 fluorescence units) and nontypeable serotype NT2 (~10000 fluorescence units) were observed (Figure 2.13). The presence of NT2 signal intensity could be attributed to cross-hybridisation because no additional HGs were detected by HGID to suggest the presence of NT2, a further confirmation of single serotype 3 carriage in the sample.

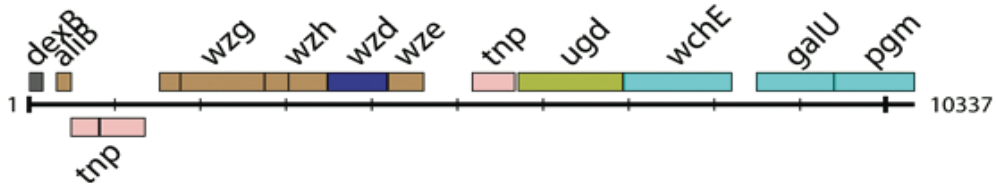


Figure 2.10 CPS locus genes for wild type pneumococcal serotype 3

Adapted from (Bentley et al., 2006). The polysaccharide capsule of serotype 3 is synthesised via synthase pathway encoded by the *wchE* gene.

2.6.4.1.1. Identifying antibiotic resistance genes in serotype 3

Using the microarray AbR probes, the presence of ten common antibiotic resistance genes in the pneumococcal genome of was tested. The signal intensity for all these genes was at the zero-baseline, which suggested the absence of antibiotic resistance genes in serotype 3 (Figure 2.14).

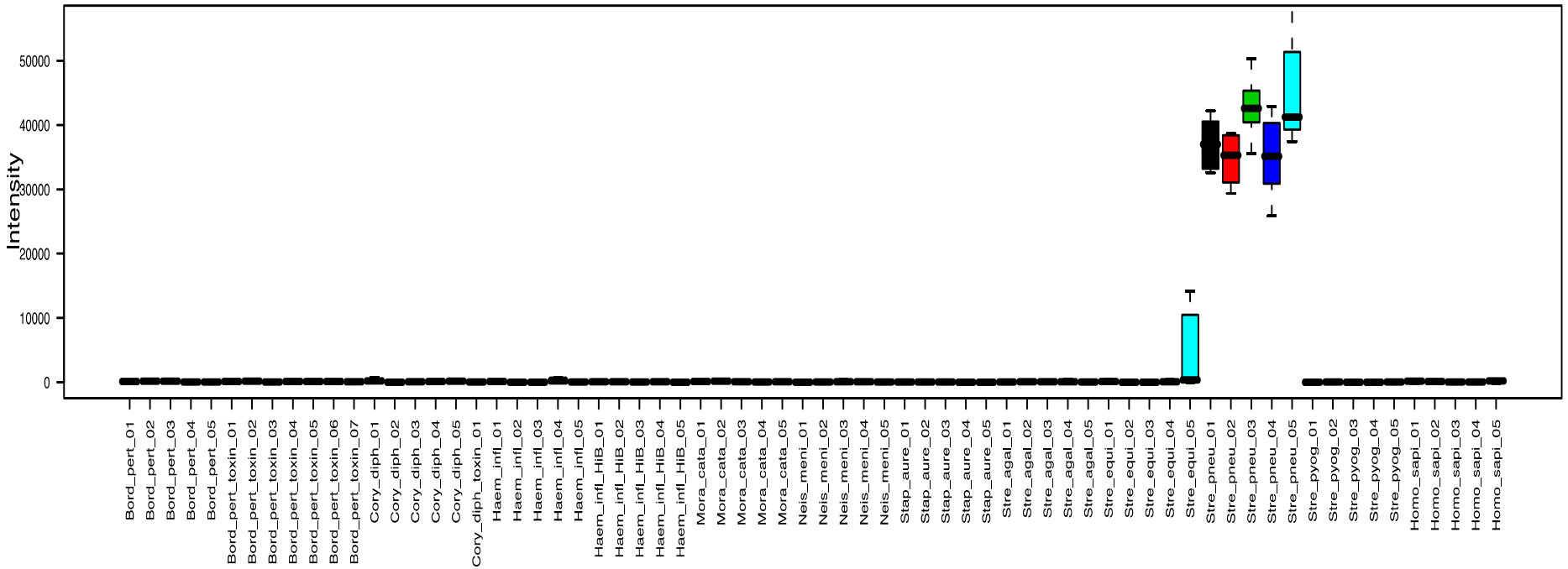


Figure 2.11 Purity of pneumococcal species

The sample (17787) was initially cultured on a pneumococcal selective media, before the DNA was extracted for analysis by microarray. The intensities of all other pathogens sit at zero baseline, while the intensity for *S. pneumoniae* is over 30000, indicating that only *S. pneumoniae* was present in the sample.

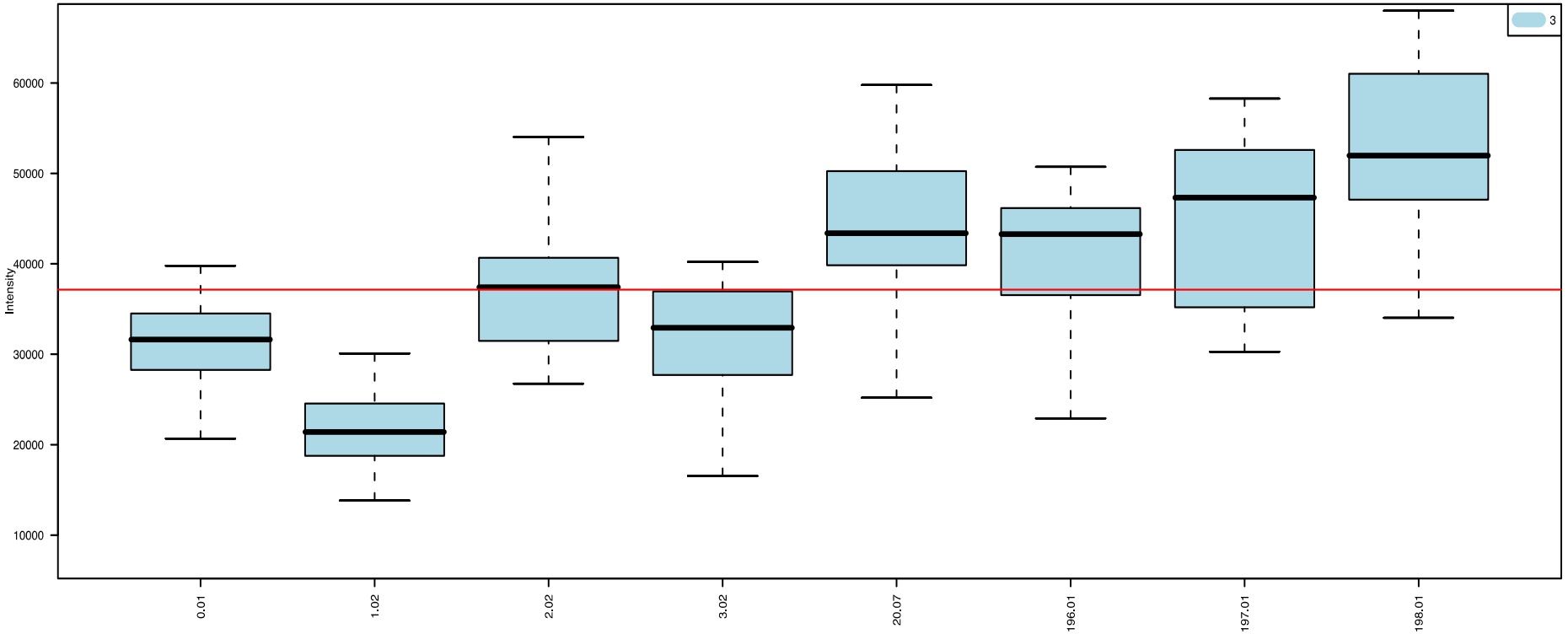


Figure 2.12 Serotype 3 homology groups

All the homology groups (horizontal axis) for the CPS genes expected for serotype 3 (sample 17787) were detected as shown (0.01 [*wzg*], 1.02 [*wzh*], 2.02 [*wzd*], 3.02 [*wze*], 20.07 [*ugd*], 196.01 [*wchE*], 197.01 [*galU*], and 198.01 [*pgm*]). The horizontal red line shows the median log intensity of the 6824 *S. pneumoniae* SpTIGR4 and R6 genome probes.

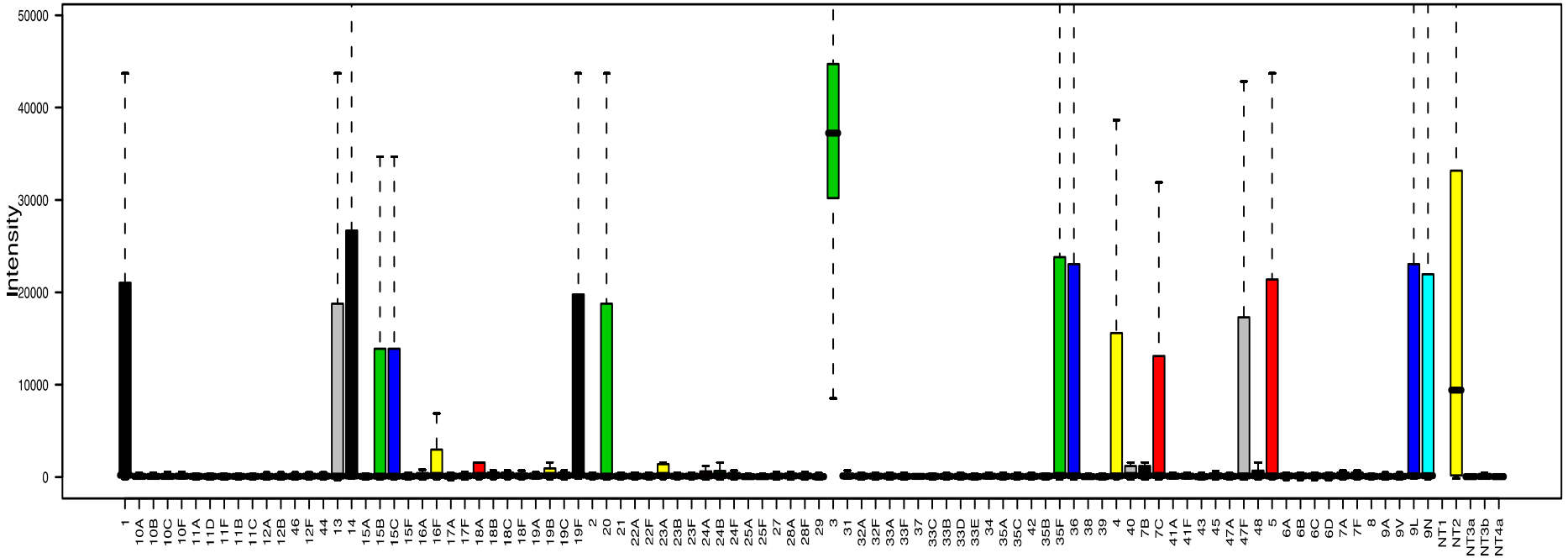


Figure 2.13 Detecting serotype 3 presence

The intensities of all other pathogens sit close to the zero baseline, while the intensity for serotype 3is over 30000, indicating that only serotypes 3 was present in the sample 17787.

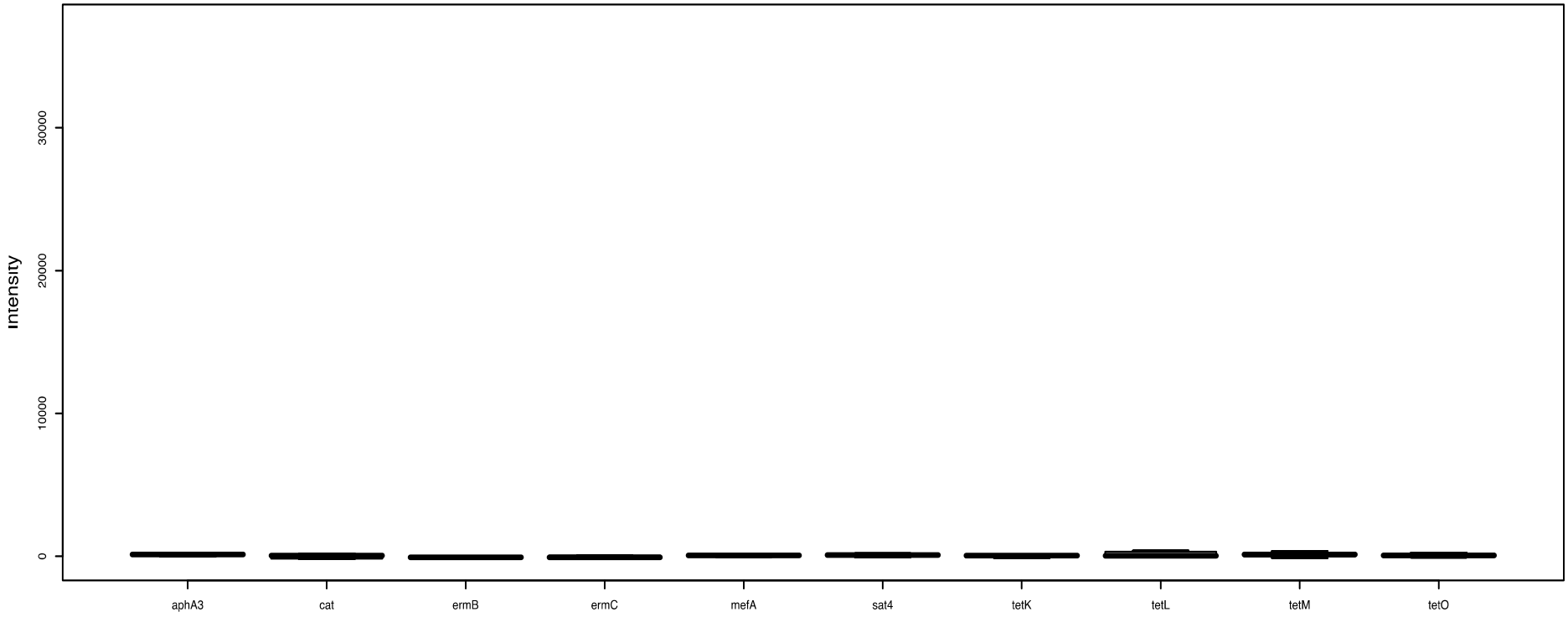


Figure 2.14 Detection of antibiotic resistance genes

The plot shows signal intensities for 10 common pneumococcal antibiotic resistance genes in sample 17787. The intensities for all the antibiotic resistance genes were at zero baseline intensity, signifying the absence of such genes from the genome of serotype 3 in sample 17787.Thus this sample would be expected to show sensitivity to treatment with the listed antibiotics.

2.6.4.2. Single carriage identification of serotype 14

Sample W01654 contained the pneumococcus only and no other pathogen was detected (Figure 2.16). The CPS locus for wild-type serotype 14 consists of 14 genes, and it is 19918 base pairs long (Figure 2.15). Serotype 14 determination by homology group presence showed fourteen HGs (Figure 2.17): 0.01 [*wzg*], 1.02 [*wzh*], 2.02 [*wzd*], 3.02 [*wze*], 5.02 [*wchA*], 7.11 [*wzx*], 26.01 [*wchJ*], 27.06 [*wchK*], 34.02 [*wchN*], 76.02 [*wchL*], 77.02 [*wchM*], 78.02 [*wciY*], and 103.02 [*wzy*], consistent with fourteen genes at the CPSlocus for serotype 14.

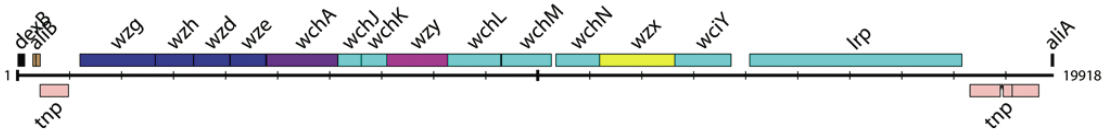


Figure 2.15 CPS locus genes for wild type serotype 14

Adapted from (Bentley et al., 2006).

Analysis by STID showed elevated signal intensities for serotype 14 (>40000 fluorescence units), serotype 15B/C (~3000 fluorescence units) and serotype 3 (20000 fluorescence units) (Figure 2.18). Homology groups for serotypes 3 and 15B/C were not detected by HGID (Figure 2.17), thus their elevated fluorescence by STID could be due to cross-hybridisation. This evidence suggests that sample W01654 contained only pneumococcal serotype 14.

2.6.4.2.1. Identifying antibiotic resistance genes in serotype 14

The signal intensity for *tetM* antibiotic resistance genes was at 40000 fluorescence units. The remaining genes were observed at the zero-baseline suggesting insensitivity to tetracycline by this particular serotype 14 (Figure 2.19).

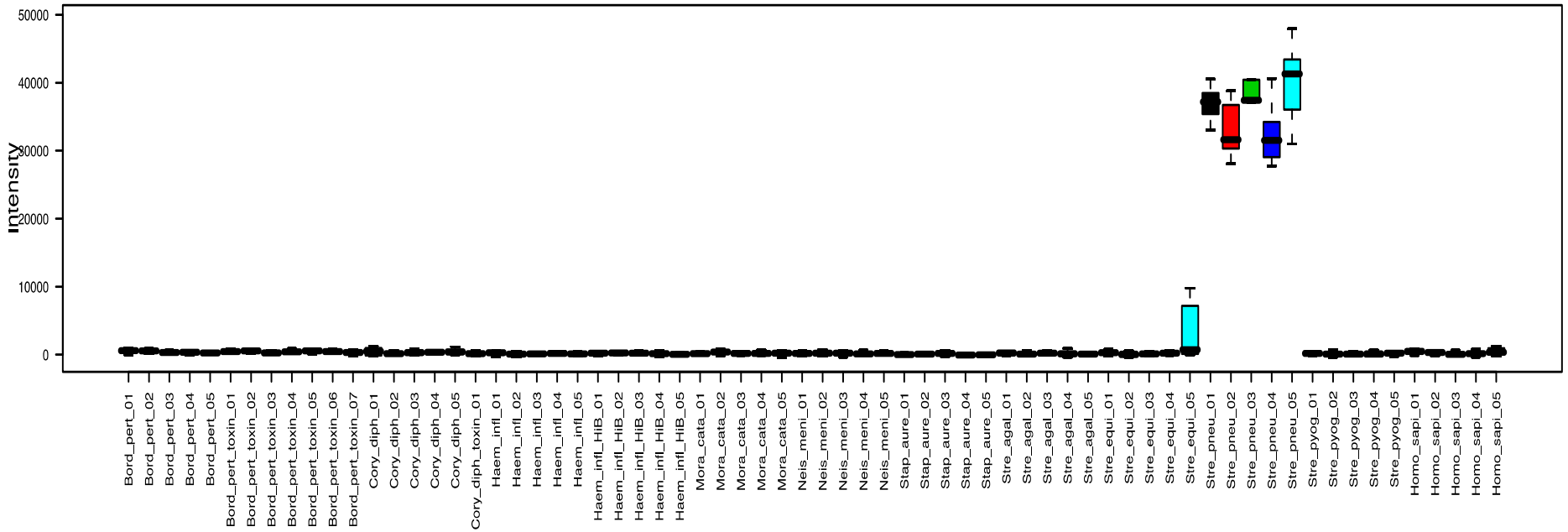


Figure 2.16 Purity of pneumococcal species

The sample (W01654) was initially cultured on a pneumococcal selective media, before the DNA was extracted for analysis by microarray. The intensities of all other pathogens sit at zero baseline, while the intensity for *S. pneumoniae* is over 30000, indicating that only *S. pneumoniae* was present in the sample.

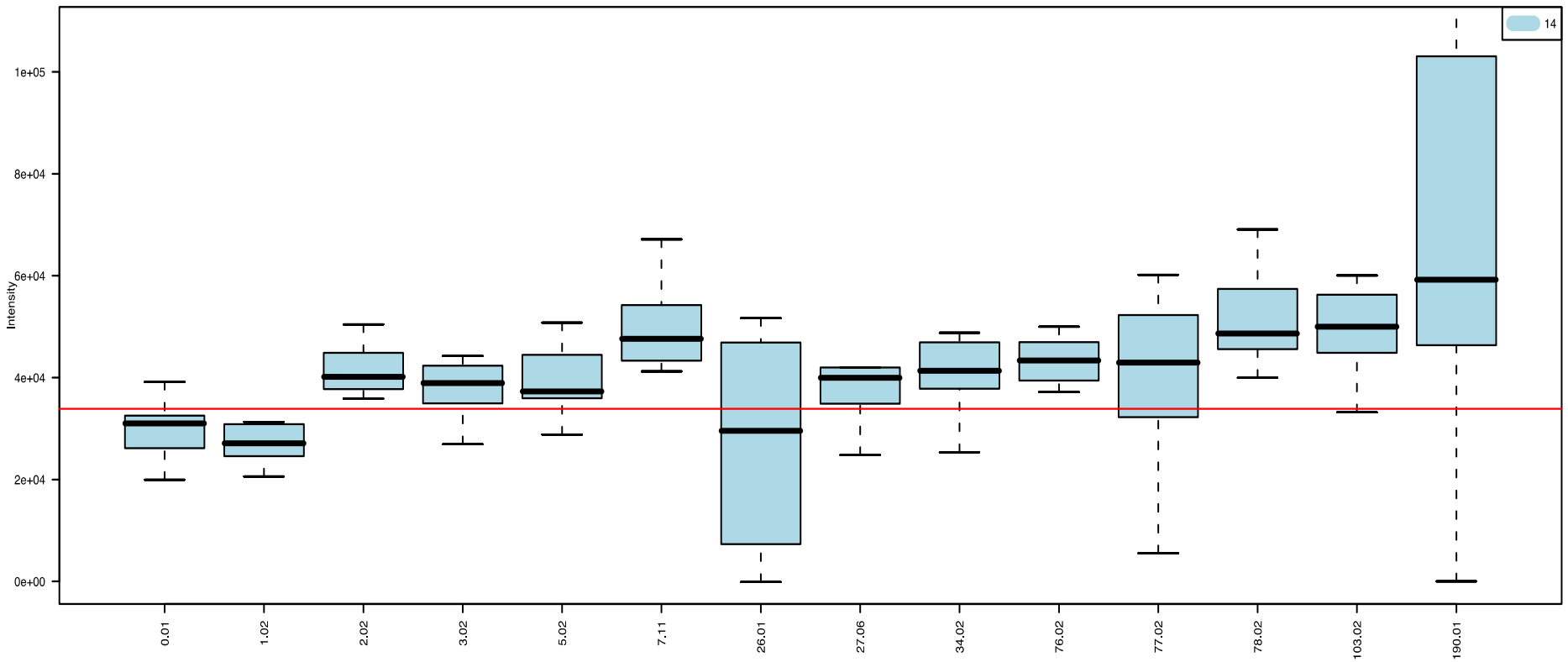


Figure 2.17 Serotype 14 homology groups

All the homology groups (horizontal axis) for the CPS genes expected for serotype 14 (sample W01654) were detected as shown. The horizontal red line shows the median log intensity of the 6824 *S. pneumoniae* SpTIGR4 and R6 genome probes.

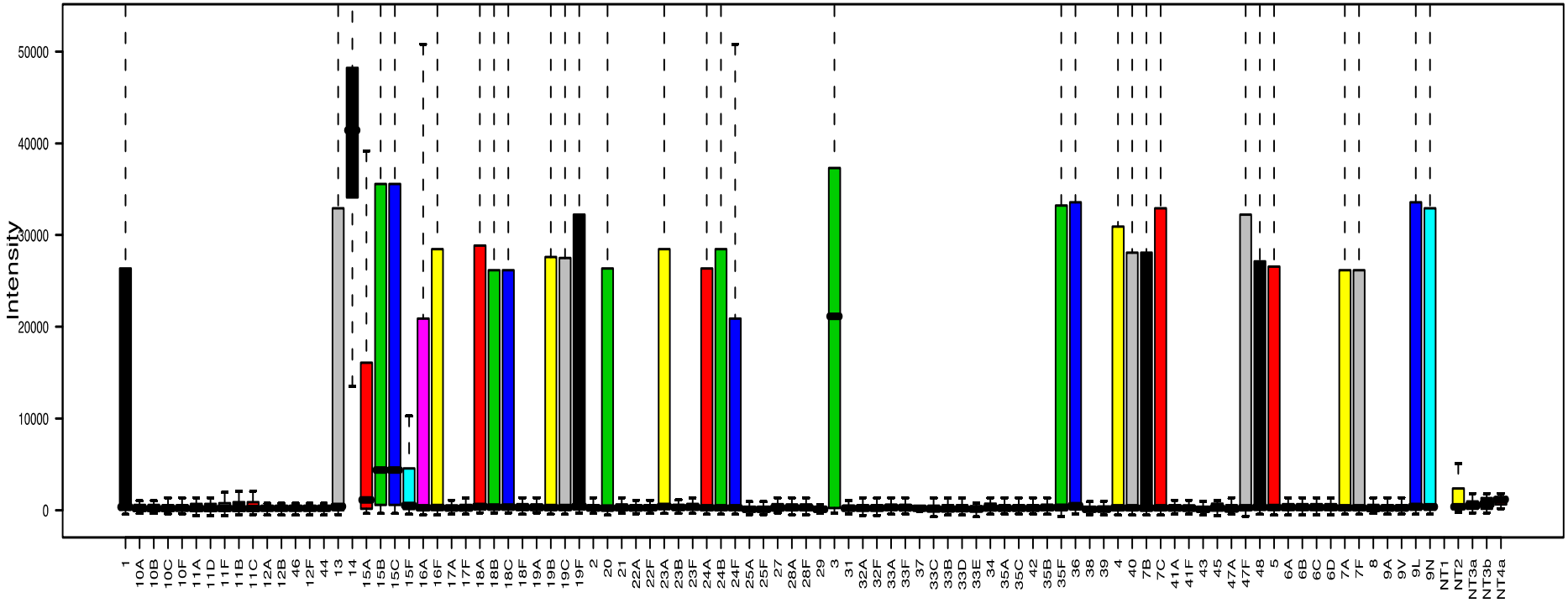


Figure 2.18 Detecting serotype 14 presence

The intensities of all other pathogens sit close to the zero baseline, while the intensity for serotype 14is over 30000, indicating that only serotypes 14 was present in the sample (W01654).

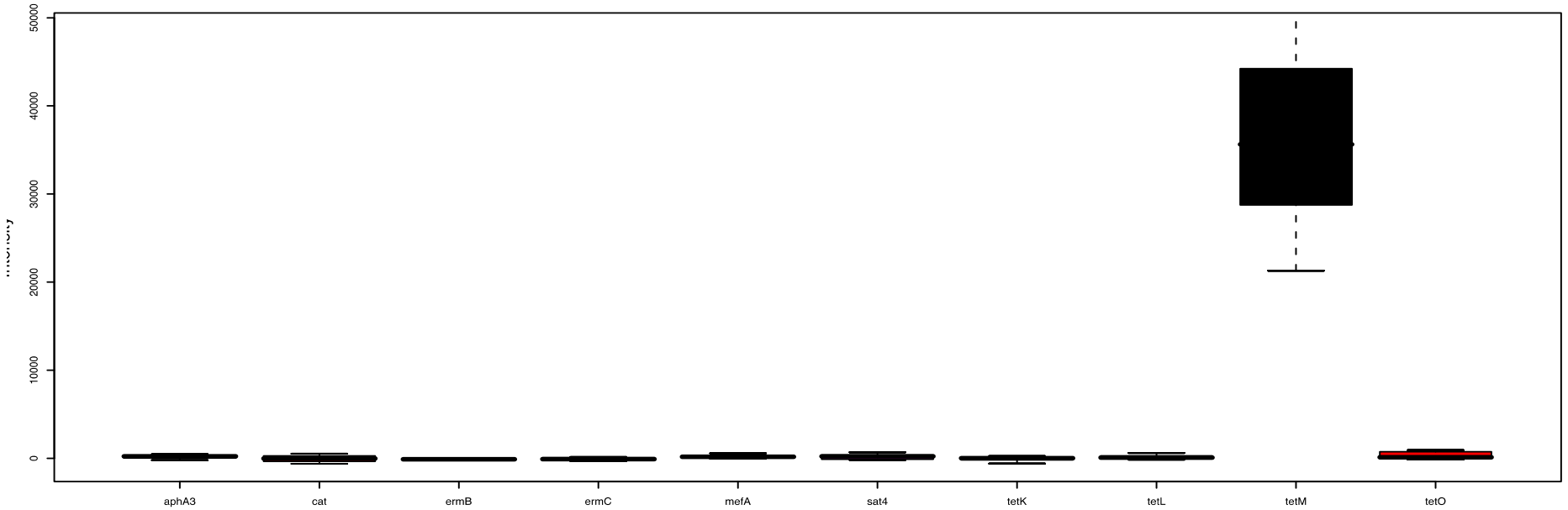


Figure 2.19 Detecting antibiotic resistance genes

The plot shows signal intensities for 10 common pneumococcal antibiotic resistance genes in sample W01654. The intensities for 9 out of the 10 antibiotic resistance genes were at zero baseline intensity, signifying the absence of such genes from the genome of serotype 14 in the sample. However, the intensity of *tetM* gene was very high at >30000 above baseline, signifying the presence of tetracycline resistance gene. Therefore, sample W01654 would be expected to show resistance to treatment with tetracycline.

2.6.4.3. Identifying multiple carriage involving serotype 3 and serotype 14

Analysis of pathogens present in sample 18262 by PathID showed elevated fluorescence for *S. pneumoniae* with all other pathogens shown at the zero-baseline fluorescence, suggesting only the pneumococcus was present (Figure 2.20). In sample 18262, multiple carriage involving serotype 3 and 14 was detected. Analysis of homology group presence by HGID, showed a combined total of eighteen HGs from serotype 3 and 14 (Figure 2.21). Four overlapping homology groups were detected in 0.01, 1.02, 2.02, and 3.02, representing *wzg*, *wzh*, *wzd* and *wze* genes respectively. The signal intensity for serotype 3 HGs was 15000 fluorescence units above baseline, which was higher than the 5000 for serotype 14, signifying that serotype 3 was the most abundant. Further analysis of serotype presence by STID, however, showed elevated signal frequencies for serotypes 3, 14 and nontypeable (NT2) (figure 2.22). The observed high frequency for NT2 by STID could be due to cross-hybridisation because it was not detected by HGID.

2.6.4.3.1. Identifying antibiotic resistance genes in sample 18262

The intensities for 9 out of the 10 antibiotic resistance genes were at baseline intensity (Figure 2.23), signifying the absence of such genes from the genomes of co-colonising serotypes 3 and 14 in sample 18262.However, the intensity of *tetM* gene was very high at >15000 fluorescence units above baseline, signifying the presence of tetracycline resistance gene. Based on the signal intensities for serotype 3 (~15000) and serotype 14 (~5000), the resistance gene could be assigned to serotype 3.However this requires further analysis by culturing the sample in a *tetM* selective media, where only the serotype carrying a *tetM* resistance gene would be expected to grow.

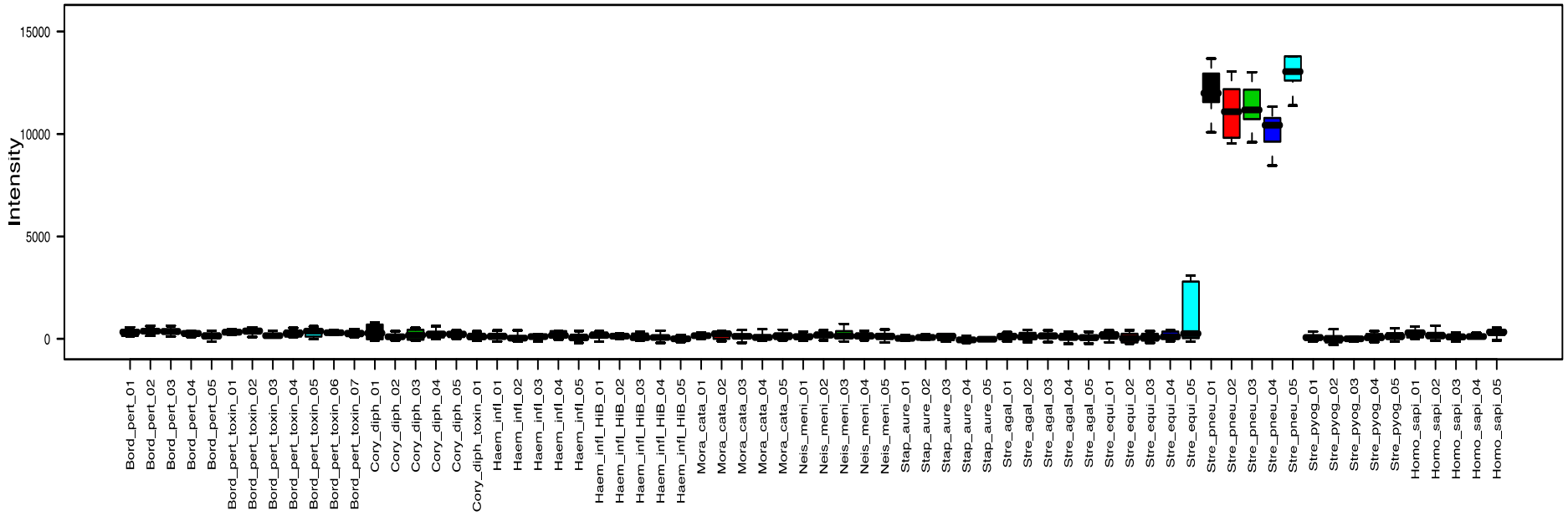


Figure 2.20 Purity of pneumococcal species

The sample (18262) was initially cultured on a pneumococcal selective media, before the DNA was extracted for analysis by microarray. The intensities of all other pathogens sit at zero baseline, while the intensity for *S. pneumoniae* is over 10000, indicating that only *S. pneumoniae* was present in the sample.

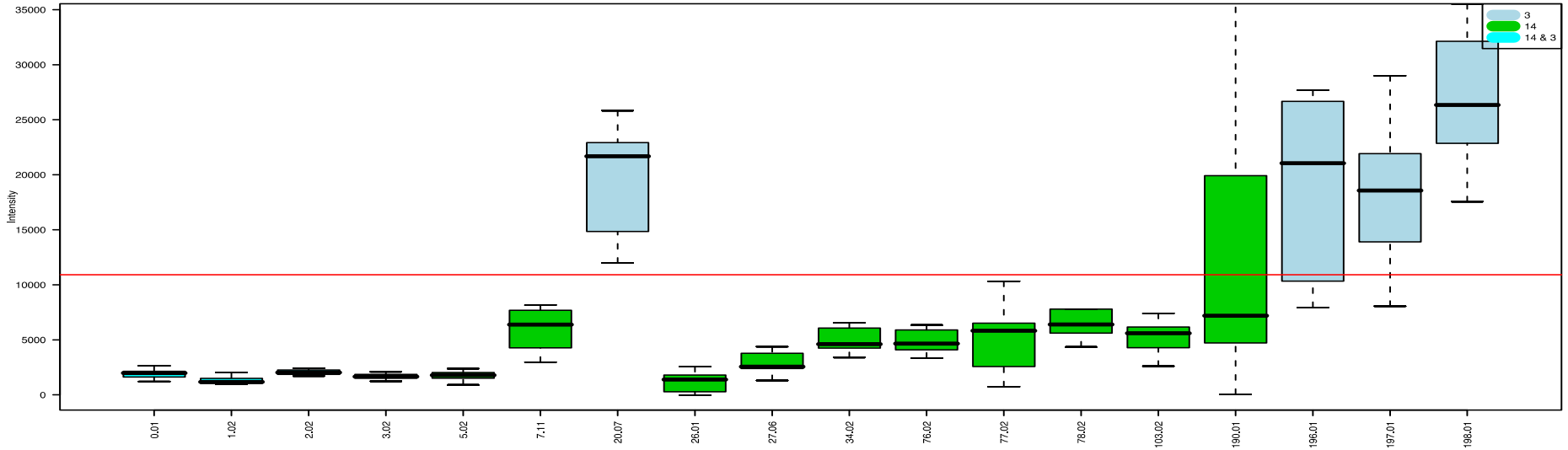


Figure 2.21 Serotype 3 and 14 homology groups

All the respective homology groups for the CPS genes expected for serotype 3 and14 were detected in sample 18262 as shown. Homology groups 0.01, 1.02, 2.02 and 3.02 are common to both serotypes 3 and 14. Homology groups shown in blue and green are unique to serotype 3 and 14 respectively. This cluster of homology groups confirms co-colonisation of serotype 3 and 14 in sample 18262.The intensity for serotype 3 HGs is 15000 above baseline, which is higher than the 5000 for serotype 14, signifying that serotype 3 was the most abundant. The horizontal red line shows the median log intensity of the 6824 *S. pneumoniae* SpTIGR4 and R6 genome probes.

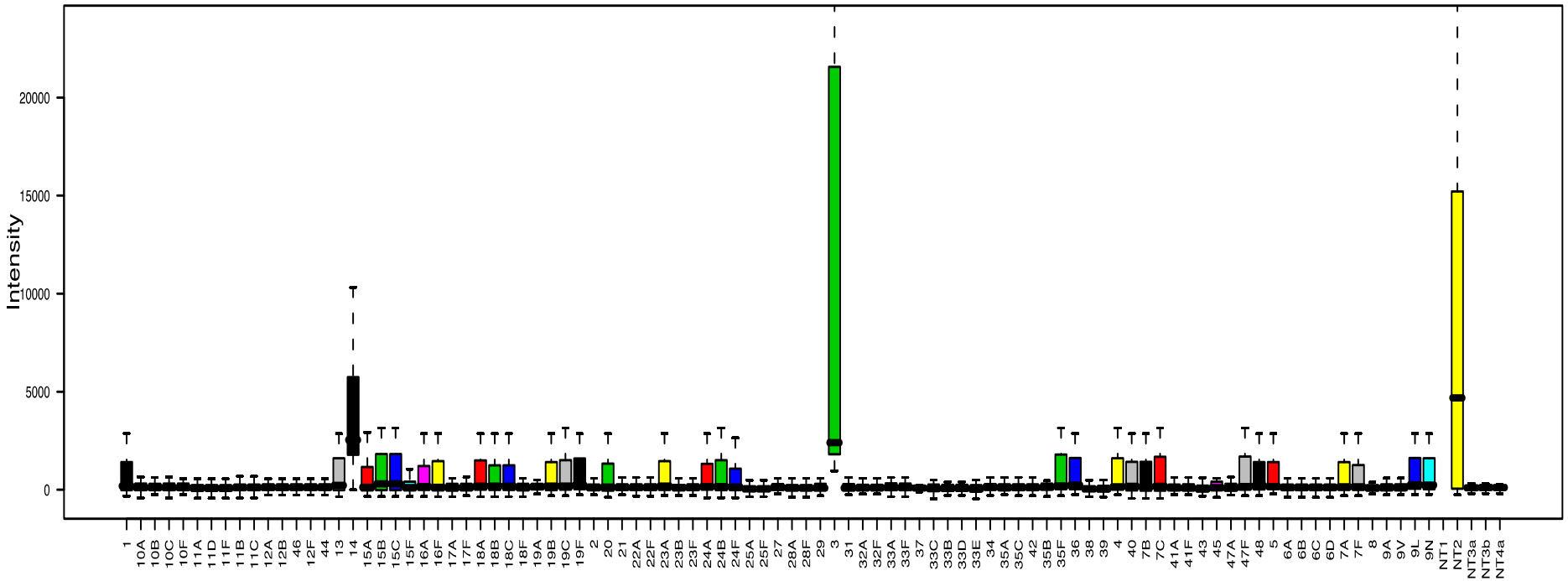


Figure 2.22 Detection of serotype 3 and serotype 14 presence

The intensities of all other pathogens sit close to the zero-baseline, while the intensities for serotype 3 and 14 are at least 5000 above the baseline, indicating co-colonisation of serotypes 3 and 14 in sample 18262.

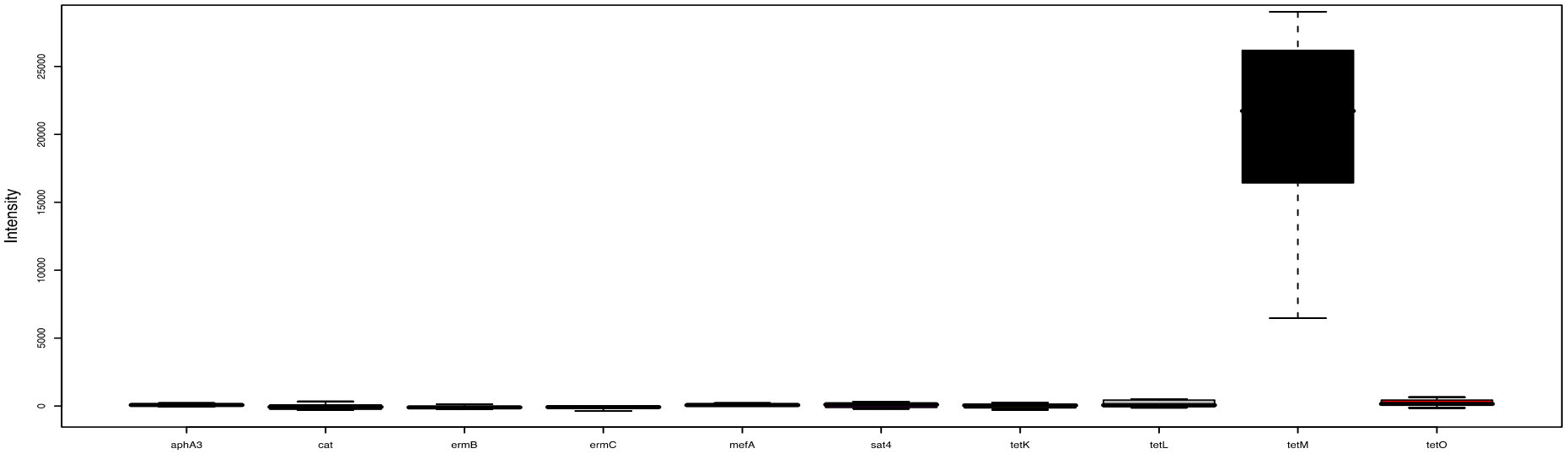


Figure 2.23 Detecting antibiotic resistance genes

The plot shows signal intensities for 10 common pneumococcal antibiotic resistance genes in sample 18262.The intensities for 9 out of the 10 antibiotic resistance genes were at zero baseline intensity, signifying the absence of such genes from the genomes of co-colonising serotypes 3 and 14 in the sample. However, the intensity of *tetM* gene was very high at >15000 above baseline, signifying the presence of tetracycline resistance gene. Based on the intensity, the resistance gene could be assigned to serotype 3, however this can be confirmed by culturing in the sample in a tetM selective media, where only the pneumococcal serotype carrying a *tetM* resistance gene would be expected to grow.

2.7. Whole genome sequencing

Whole genome sequencing was performed on Illumina HiSeq platform (CA, USA). Prior to sequencing, the DNA sample was prepared into a sequencing library by random fragmentation of DNA sample into pieces each about 200 bases long (Figure 2.24). Custom made nucleic acid adapters were attached to the blunt ends of the sheared DNA fragments, and then flowed through a solid surface containing immobilised template nucleic acid fragments that are complementary to the adapters. The bound fragments undergo a solid phase bridge PCR amplification to create clonal clusters of single stranded DNA molecules, containing identical DNA sequences. The amplified fragments are denatured to obtain single stranded DNA templates attached to the solid phase.

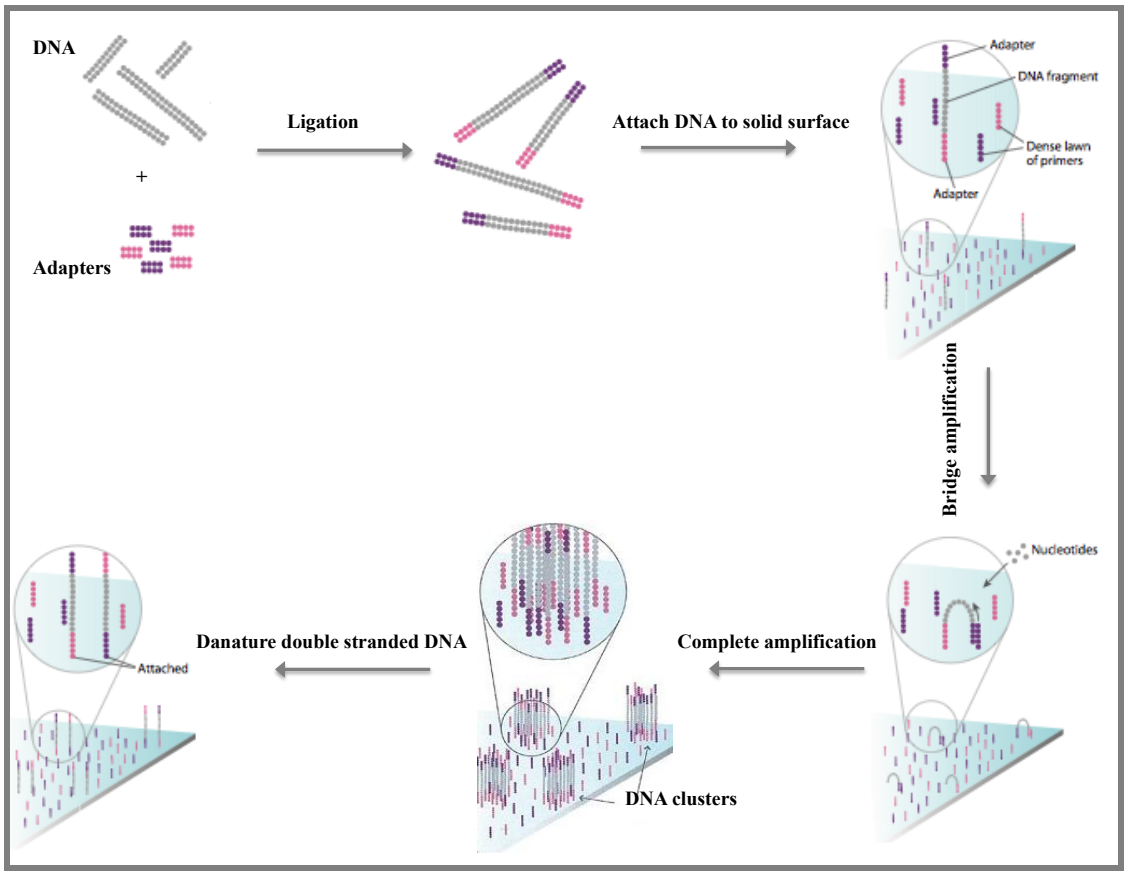


Figure 2.24 Preparation of DNA library for Illumina sequencing

Randomly fragmented DNA is attached (ligation) to adapters at both ends of the fragment. The single stranded fragments are bound to the solid phase and then subjected to bridge PCR amplification. After complete amplification, the DNA is denatured to obtain single strands prior to sequencing (Mardis, 2008).

The HiSeq illumina platform uses sequencing by synthesis technology, similar to Sanger sequencing except for the usage of modified dNTPs. The modified dNTPs contain a terminator molecule at $3'$ -OH ends to prevent further polymerisation. This ensures that only a single nucleotide is added to each growing DNA strand by the DNA polymerase enzyme. The sequencing reaction is initiated by adding primers to the cluster strand generated by bridge amplification followed by fluorescently labelled modified dNTPs and DNA polymerase. The excess nucleotides and DNA

polymerase molecules are washed away; a scan buffer is added to the flow cell to capture the image for each lane of the flow cell. Following image capture, the fluorescent labels and the 3′-OH blocking groups are removed to prepare the cluster strands for another cycle of fluorescent nucleotide incorporation (Mardis, 2008).

2.8. Whole genome assembly

The sequences were assembled using Velvet, a de novo short sequence read assembly using de Bruijn graphs (Zerbino and Birney, 2008) (Figure 2.25). The sequence reads for CPS locus variants (Chapter 5) were mapped against respective reference sequences (Bentley et al., 2006) using Burrows Wheeler Aligner (Li and Durbin, 2009). The SNPS were called using SAMTOOLS, VCFTOOLS and BCFTOOLS (Li et al., 2009). The alignment was generated based only on single nucleotide polymorphisms (SNPs) from the whole CPSlocus gene sequence. The sequence reads for sequential serotypes were mapped against reference genome *S. pneumoniae* ATCC 700669 (Accession number FM211187), and the alignment was generated based on whole genome sequence SNPS. The size for most of the contigs obtained after assembly ranged from ~20000 to ~60000 base pairs (Appendix 3a), while the cumulative length form the largest contig to the smallest contig with at least 50% of the total length (N50 contig number) ranged from ~50000 to ~100000 base pairs (Appendix 3b).

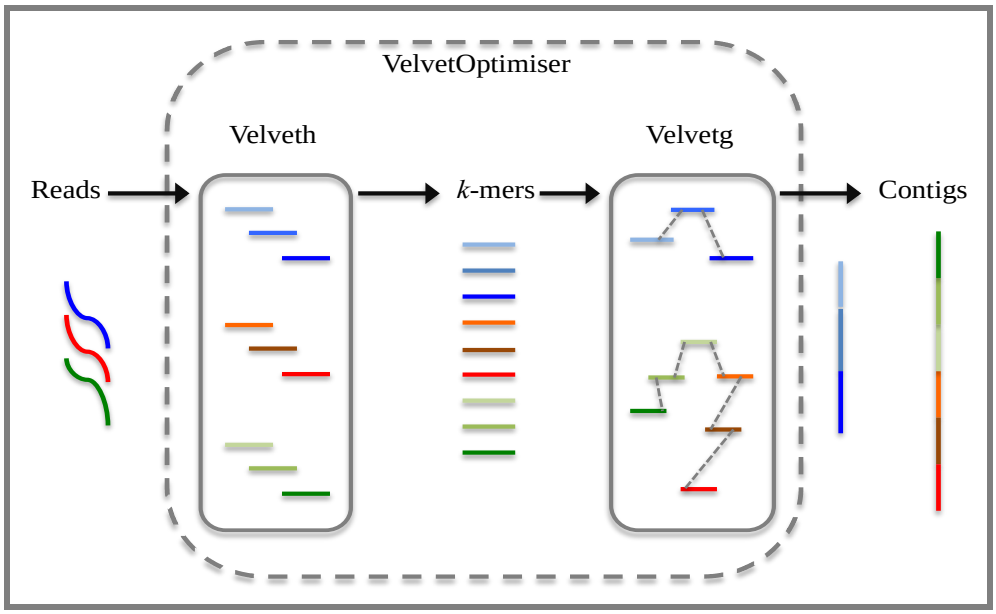


Figure 2.25 Assembly of sequence reads using Velv*et* and VelvetOptimis*er*

This is performed in two steps, (1) *Velveth* converts reads to very short reads (*k*-mers) using a hash table, which is a data structure used to implement an associative array (key/value pairs), and (2) *Velvetg* assembles overlapping *k*-mers into contigs via a de Bruijn graph. A de Bruijn graph is described as a compact representation based on short reads (Zerbino and Birney, 2008).

2.8.1. Phylogenetic analysis

By definition, phylogeny refers to the evolutionary history of a group of organisms. A phylogenetic tree is often generated to depict an interpretation of phylogeny (Soltis and Soltis, 2003). In genetics, evolutionary trees are often estimated from DNA or RNA sequences. A determination of the confidence levels for the generated trees is by the application of bootstrapping technique. Bootstrapping estimates confidence levels of inferred relationships by resampling the original data matrix with replacement of the characters, to produce replicate bootstrap data sets. This process is repeated many times and phylogenies are reconstructed each time. Once

bootstrapping is complete, a consensus tree is constructed from the optimal tree in each bootstrap sample. The bootstrap support for any internal branch is the number of times it was recovered during the bootstrapping procedure; often referred to as the bootstrap value, bootstrap percentage or bootstrap *p*-value (Efron et al., 1996, Soltis and Soltis, 2003). In this study, the phylogeny of sequenced genomes was determined using the alignment of SNPs from each genomic sequence with RAxML (Stamatakis et al., 2005) by performing 100 bootstraps. The genome of serotype 23F *S. pneumoniae* ATCC 700669 (Accession number FM211187) was used as a reference genome for inferring phylogenetic relationships. However, for capsular polysaccharide locus variants of serotypes 6B, 19A, and 20, the reference sequences of 6B (GenBank® accession no. CR931639), 19A (GenBank® accession no. CR931675) and 20 (GenBank® accession no. CR931679) were used respectively. The genetic relatedness of these genomes was further established by Artemis, ACT®(Rutherford et al., 2000, Carver et al., 2005) or Easyfig (Sullivan et al., 2011).

2.8.2. Pneumococcal genotyping by MLST

Multi-Locus Sequence Typing (http://www.spneumoniae.mlst.net) is a molecular typing technique for determining the population structure of bacterial species and other related organisms (Maiden et al., 1998). For pneumococcal strains, MLST typing is based on characterisation of seven housekeeping genes namely: *aroE, gdh, gki, recP, spi, xpt* and *ddl* (Enright and Spratt, 1998). In each locus, alleles are assigned an arbitrary number based on sequence variations (Maiden et al., 1998). In this study, MLST was performed using an in-house script developed by Dr Nicholas J. Croucher (get_st.shmlst_genes.embl forward_reads.fastq) at the Wellcome Trust Sanger Institute in Cambridge, UK. Briefly, the Illumina short sequencing reads were

mapped against the sequences of the seven loci used for sequence typing in order to determine the MLST sequences for each of the strains used (Croucher et al., 2011). The sequences for each isolate were then analysed using MLST (http://www.spneumoniae.mlst.net) in order to determine their allelic profiles and corresponding sequence types (STs) (Aanensen and Spratt, 2005).

2.8.3. Genetic recombination

Genetic recombination involves acquisition of genetic elements from other microorganisms through transformation, transduction or conjugative transfer (Thomas and Nielsen, 2005). In this study, genetic recombination was assessed in Malawian carriage serotypes, using an in-house script for identifying genetic recombination events (gubbins.py -a input.aln -p output_prefix -t phylogeny.tre) developed by Dr Simon R. Harris at the Wellcome Trust Sanger Institute in Cambridge as reported previously (Croucher et al., 2011). Briefly, ancestral sequences were reconstructed onto each node of the phylogenetic tree using a package of phylogenetic analysis programs called Phylogenetic Analysis using Maximum Likelihood (PAML) developed by Ziheng Yang (Yang, 1997). From the ancestral sequences, SNPs were reconstructed onto the branches of the phylogeny. Recombination events were identified using a sliding window approach along the sequences. Genomic regions with high SNP density within the genomes were inferred to represent occurrences of the genetic recombination events. This is because mutations occur spontaneously in the genomes due to errors in DNA replication as such these are uniformly distributed across the whole genome. The implication is that sections of the genome with unusually high SNP or polymorphism density would indicate nonrandom mutations, and therefore recombination.

2.9. The potential of 6B variants to colonise and cause death in mice

2.9.1. Bacterial strains

The serotype 6B CPS locus variants used were obtained from the nasopharynx of Malawian children. For comparison, previously characterised Malawian invasive 6B serotypes (Everett et al., 2012) were included. Pneumococcal strains were routinely cultured on blood agar baseplates containing 5% (v/v) horse blood or in brain heart infusion (BHI) broth (BHI; Oxoid, Basingstoke, UK) containing 20% (vol/vol) fetal bovine serum (FBS; Gibco, Paisley, UK).

2.9.2. Preparation of the challenge dose

BHI broth (10 ml/tube/strain) was inoculated with four to five colonies taken from a fresh culture plate of each respective 6B *S. pneumoniae* strain and incubated overnight at 37 °C. Bacteria were harvested by centrifugation (18,000*g)* and resuspended in 1 ml of fresh serum broth (20% FBS in BHI), and the pellet suspension was diluted with fresh serum broth to give an optical density at 500 nm of 0.7. The culture was then incubated at 37 °C for a further 4 to 5 hours to reach an optical density of 1.6 at 500 nm. Aliquots of this culture were stored at -80 °C. Viable colony counts of thawed aliquots were performed as described below in duplicate on BAB 5% (v/v) horse blood. Right before administration into mice, the suspension was thawed slowly at room temperature, and bacteria were harvested by centrifugation before re-suspension and dilution at 2.10^7 CFU/ml in sterile phosphate-buffered saline (PBS).

2.9.3. Viable cell counting

Serial dilutions of 20µl samples were performed in sterile 96-well microtiter plates (GIBCO, Paisley, UK) containing 180µl/well of sterile PBS per well and serial dilutions were made to 10^6 with each strain. Dried 5% horse blood (v/v) BAB plates were marked into six sectors, and three 20µl aliquots of each dilution were plated into each sector. Once dry, plates were incubated at 37 °C overnight. Colonies were counted in sectors containing a measurable number of colonies (between 30 and 300).

2.9.4. Outbred mouse strain

MF1 mice (n=10 mice per bacterial isolate) were obtained from Charles River Laboratories (Kent, UK). Mice were obtained at 8 weeks old and infected when more than 9 weeks old. After arrival, all animals were kept for a minimum of 7 days to acclimatise. Following infection, mice were housed in an isolator.

2.9.5. Intranasal challenge of mice and experimental endpoint

Mice were anaesthetised with a mixture of O_2 and isofluorane, and each mouse was infected intranasally with 10^6 CFU *S. pneumoniae*in a volume of 50µl. The numbers of bacteria in the blood were monitored at 6, 12, 24, 48 and 72 hours post-challenge; up to 50µl of blood was taken from the tail vein, and viable counts were performed. Following infection, mice were closely monitored for the visual development of symptoms, as described previously (Morton and Griffiths, 1985). Mice were closely monitored for clinical signs of disease and were culled when severely lethargic. The time point the animal was scored as severely lethargic was recorded, and the animal was then immediately culled by CO_2 euthanasia. Data were represented as Kaplan-

Meier survival curves. Mice were monitored up to 7 days (or 168 hours), at which point the experiment was ended. Seven days was the optimised time period to monitor the impact of intranasal challenge on mice. Mice that were alive at day 7 were considered to have survived the pneumococcal challenge. Nasopharynx and lungs were removed into 3 and 5 ml of sterile PBS, respectively. Whole blood samples were collected by cardiac puncture into sterile heparinized tubes. The tissues were then homogenized using an IKA T10 basic Ultra-Turrax® Homogenizer system 115VAC (IKA, Staufen, Germany). Viable counts in tissue homogenates and in blood samples were performed as described above. Monitoring of the challenged mice was done in a blinded manner i.e mouse cages were moved around and group ID was hidden by the laboratory staff not involved in the experiments during each observation.

2.10. Investigating nasopharyngeal microbiota in children and adults

2.10.1. The 16S rRNA bacterial classification by 454 pyrosequencing

The microbial diversity in the nasopahrynx of adults and children was investigated using 454 sequencing (Chapter 1, Section 1.11.1) of the PCR amplified V3-V5 region of the 16S rRNA. The principle of 454 is through sequencing-by-synthesis (Rothberg and Leamon, 2008)

2.10.2. DNA Extraction

DNA was extracted directly from samples in STGG using a FastDNA spin kit (MP Biomedicals, USA) for soil, following the manufacturers' protocol. Nasopharyngeal swabs in STGG were vortexed briefly. An aliquot of the STGG solution (200μl) was transferred into the tube containing a lysing matrix of beads with mixed sizes.

Sodium phosphate buffer (978µl) and MT buffer (122µl) were added to the tube and sealed securely with parafilm wrap. The MT buffer is composed of molecular grade water, Ethylene diamine tetra-acetic acid disodium salt dehydrate, sodium dodecyl sulphate, polyvinlypyrolidone and proprietary inorganic salts (http://www.mbio.com). Sodium phosphate buffer and MT buffer function to protect and solubilize nucleic acids and proteins during cell lysis, and also minimise RNA contamination. The mixture was homogenized in Ribolyser cell disrupter (Hybaid) for 30 seconds at the speed of 5.5, and then centrifuged at 14, 000g for 5 minutes. The supernatant was transferred to a clean 2ml eppendorf tube. The protein precipitation solution (PPS) (250µl) was transferred to the eppendorf and mixed manually by inverting the tube several times. The mixture was centrifuged as before to pellet the precipitate. The supernatant was transferred to a clean 15ml falcon tube. The binding silica matrix suspension (1 ml) was added to the supernatant in the falcon tube. The samples in the tubes were mixed on the grant-Bio shaker at 14000rpm for 2 minutes, to allow binding of DNA to the matrix. The tubes were placed in a rack for 3-5 minutes to allow the silica matrix to settle. The supernatant (500µl) was removed from the top of the tube, without disturbing the settled matrix. The remaining sample/matrix mixture was re-suspended by vortexing briefly. The suspension (600µl) was transferred into the top of the spin filter tube and centrifuged at 14,000g for 1 minute. The silica matrix was held on the column while the filtrate (liquid) from the catch tube was removed and discarded. The process was repeated until the entire sample in the falcon tube was filtered. The silica matrix on the spin filter membrane was washed by carefully re-suspending in SEWS-M (500µl), a wash solution composed of water and tris-hydrochloride, and then centrifuged at 14000g for 1 minute. The liquid from the catch tube was removed and discarded. A further

centrifugation was performed at 14000g for 2 minutes to dry the filter. The filter was removed from the catch tube and placed in a clean new tube. The filter was allowed to dry by incubation at 30 °Con a heating block. The ultra-pure water (DES) (100µl) was added to the filter and the silica pellet was carefully re-suspended with a pipette tip (being careful to avoid perforating the filter with the pipette tip). The tube was then incubated at 55 °C for 5 minutes to allow efficient dissolution of the matrix bound Dante tubes were centrifuged at 14,000g for 1 minute to transfer the eluted DNA into the clean tube. For a final wash, the eluted DNA was added into a micron 30kDa filter containing 400µl of 1x Tris-EDTA (TE) buffer. The solution was subjected to centrifugation at 14,000g for 11 minutes, discarding the flow through. The micron filter was inverted into a clean 1.5ml tube and centrifuged at 1000 rpm for 1 minute to elute the washed DNA (30µl). The DNA was stored at -20 °C.

2.10.3. PCR amplification of 16S rRNA

The V3-V5 region (~450bp) of the16S rRNA genes was PCR amplified from the template DNA using universal forward 338F primer and a reverse 926R primer, and the Accuprime ™ Taq DNA Polymerase High Fidelity kit (Invitrogen, Germany). The reverse primer (Appendix 4) contained an adaptor sequence for 454 sequencing, a unique barcode sequence for identification of samples in the mixture, and a universal primer sequence for the amplification of the 16S rRNA gene (Figure 2.26)

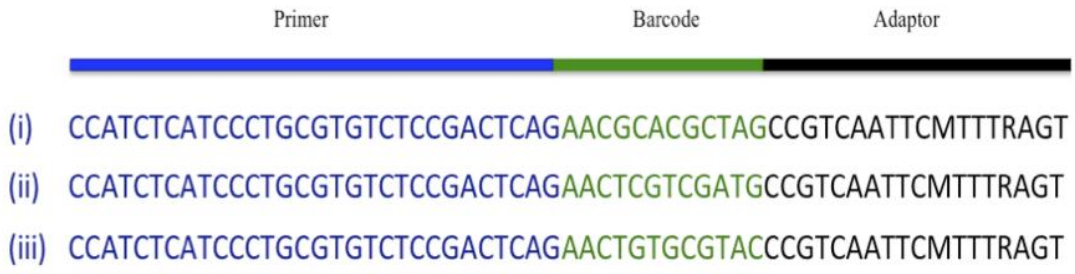


Figure 2.26 Structure of the barcoded reverse 926R primer

Primers for three different samples (i), (ii), and (iii) are shown. The primer and the adaptor sequences are similar for all the samples; however, the barcode is unique to a particular sample. The primer is used for 16S rRNA gene amplification (Parameswaran et al., 2007).

The binding positions for 338F and 926R primers on the 16S rRNA gene are 338-355 and 926-907 respectively (Mao et al., 2012). For the amplification of the V3-V5 16S rRNA region in this study, the forward (338F) and reverse primer (926R) targets were at positions 350-400bp (v3f) and 950-900bp (v5r) respectively (Figure 2.27).

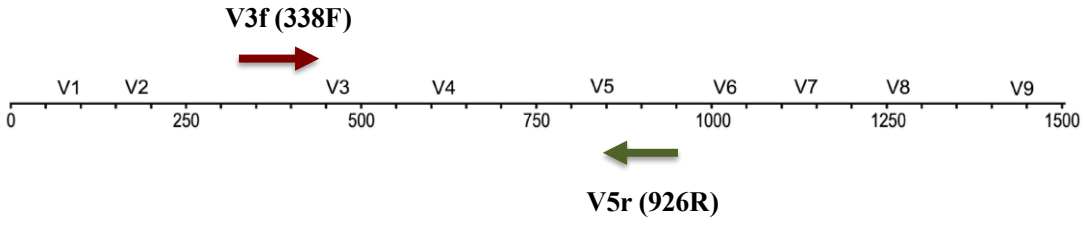


Figure 2.27 PCR primer binding positions on 16S rRNA gene

Adapted from (Claesson et al., 2010).

For each sample, 4-6 PCR reactions were performed. The number of reactions was dependent on the yield of the PCR product. The total volume of the reaction mixture

per sample was 20µl, mixed as follows: buffer II (2µl), water (15.52µl), forward primer, (0.2µl), barcoded reverse primer (0.2µl), Accuprime Taq (0.08µl), template DNA (2µl). For the negative control reaction, water or TE buffer instead of DNA was used. PCR reactions were performed in a 96-well plate, loading 20µl of the reaction mixture in each well. The PCR run conditions were as follows: 94 ºC for 2mins; (94 ºC for 30secs, 53 ºC for 30secs, 68 ºC for 2mins) x30. The 30 PCR cycle number was used to increase the product yield because of the low yield of template Dante replicate PCR products for each sample were pooled into a new 1.5ml tube. The following was added to precipitate the pooled products: NaCl (1M, 0.3 volumes relative to the pooled sample volume) and ice-cold ethanol (100%, 2 volumes relative to volume of the new combined (sample + NaCl). The solution was mixed by inversion and left overnight at -20 °C. This was followed by centrifugation at 4 °C at 16,000rpm for 30mins. The supernatant was gently removed by pipetting and the pellet was washed with ice-cold ethanol (70%, 600µl) by centrifugation as before. The supernatant was gently removed and the tubes were left to air dry for 30 – 60mins. The pellet was re-suspended in TE buffer (20µl) and left overnight at 4 °C.

2.10.4. Determining the concentration and size of the 16S PCR product

The concentration of the 16Samplicon was determined using Qubit® 2.0 fluorometer (Invitrogen, Life technologies, Austria) following the manufacturers protocol. The high sensitivity setting for the qubit was chosen because of the low DNA yield from 16S PCR amplification. The size of the PCR product (450bp) was confirmed by running on a 1% agarose gel (Figure 2.28). A mixture of the pooled 16S PCR products from each sample was prepared, by pipetting an appropriate volume to contain equal amount of DNA in each sample (equimolar mixture). The total

concentration of DNA in the pooled equimolar mixture was ~45ng/µl. The total volume of the equimolar mixture was approximately 2.6ml.

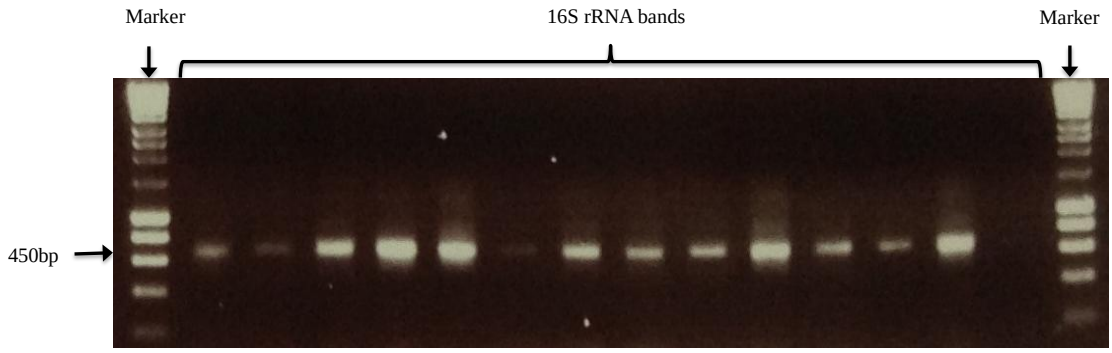


Figure 2.28 Gel image of PCR amplified 16S rRNA

The 450bp band corresponds to 16S rRNA gene. The samples were run on a 1% agarose gel, using ethidium bromide as a fluorescent dye.

2.10.5. Purification of the amplicon mixture

The equimolar mix was run through a gel clean up with the Wizard SV Gel and PCR Clean-Up System (Promega, Madison, USA) following the manufacturers protocol. This was necessary to remove unreacted PCR reagents as well as possible primer-dimers. Primer-dimers are non-specific PCR products resulting from template-independent primer interactions (Rychlik, 1995). The equimolar mixture was loaded onto low melt agarose gel and run with Tris-acetate-EDTA (TAE) buffer in the presence of ethidium bromide. The DNA in the gel was visualised and photographed using log-wavelength UV lamp. To reduce damage to the DNA, the gel was irradiated for not more than 4 seconds. The PCR product was excised using a clean scalpel, excluding the primer dimer band (Figure 2.29). The gel slice was transferred to the pre-weighed 1.5ml micro-centrifuge tube and the total weight was recorded.

The weight of the gel slice was determined as the difference between the total weight (gel slice + micro-centrifuge tube) and the weight of the micro-centrifuge tube. The membrane binding solution was added to micro-centrifuge tube at a ratio of 10µl of solution per 10mg of agarose gel slice. The mixture was incubated at 65 °C, vortexing every few minutes, until the gel slice was completely dissolved. The tube was centrifuged briefly to ensure the contents were collected at the bottom of the tube. The DNA was recovered from the dissolved gel slice by column centrifugation. The dissolved gel mixture was transferred to the SV Minicolumn (Promega, Madison, USA) assembly and incubated for one minute at room temperature, followed by centrifugation at 16,000xg for a minute. The column was removed from the assembly and flow through in the collection tube was discarded. The column was placed back on the collection tube. The column was washed by adding 700µl of membrane wash solution, and then centrifuged at 16,000xg for one minute. The collection tube was emptied as before. The wash step was repeated with 500µl of wash solution and centrifuged as before for 5minutes. The collection tube was emptied and centrifugation was performed for a further one minute with the column lid open to ensure evaporation of any residual ethanol. The SV Minicolumn was carefully transferred to a clean 1.5ml micro-centrifuge tube. Nuclease-free water (50µl) was applied directly to the center of the column, avoiding touching the membrane with the pipette tip. After incubating at room temperature for one minute, centrifugation was performed for one minute at 16,000xg.The column was discarded and the micro-centrifuge tube containing the eluted DNA was stored at -20 °C.

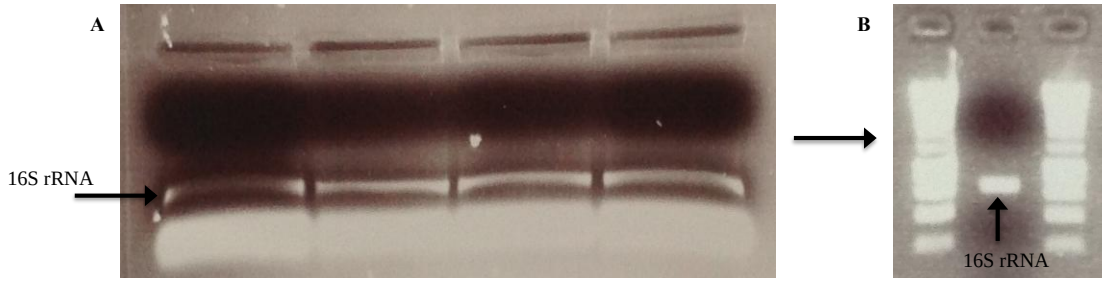


Figure 2.29 Purification of PCR amplified 16S rRNA

Purification was performed on a low melt gel (A) using Wizard SV Gel and PCR Clean-Up System (Promega, Madison, USA). The 16S rRNA band indicated by the arrow in (A) was cut out using a clean scalpel and re-extracted from the gel. The purity of the 16S rRNA was confirmed by running on 1% agarose gel (B).

2.10.6. Sequencing of the 16S PCR amplicon

The 16S PCR amplicons were sequenced using 454 sequencing (Margulies et al., 2005), on the Genome Sequencer FLX^{TM} System (Roche, Basel, Switzerland). The V3-V5 amplicons were sequenced in reverse from the 3/ end of the amplicon back to the 338f primer, because the barcode was on the reverse 926R primer. The average sequence read length obtained was 528 base pairs.

2.10.7. Analysis of 454 16S DNA (16S rRNA) sequences in Mothur

In this study, quality 16S DNA sequences were obtained using an open-source mothur software package (v. 1.30.2) (Schloss et al., 2009). Mothur is a single software command-line computer program that is built upon other pre-existing tools for analysing sequence data from microbial communities. Basically, mothur can be used trim, screen, and align sequences; distance matrix-based sequence clustering, assign representative sequence reads to operational taxonomic units (OTUs); and to

provide an estimation of within- (α-diversity) and between- (β-diversity) sample diversity (Schloss et al., 2009) (Figure 2.31). Each sequence read was identified by its barcode followed by the removal of primer sequences. The sequences were trimmed using a moving window size of 50 base pairs with the minimum average quality score in that window of 35. Therefore trimming was done to the end of the last window with an average of over 35. Any sequence containing a homoploymer sequence stretch of more than 8 nucleotides and an ambiguous base call "N" was removed. Sequence reads shorter than 400 nucleotides were also removed. Since sequencing was conducted in reverse, the reverse complement of each sequence was obtained. The sequences were then aligned using reference silva alignment template, a database of 16S sequences (Quast et al., 2013). The aligned sequences were screened to maximise the number of sequences with the longest overlapping span. Generation of chimeric sequences is very common during co-amplification of closely related sequences such as 16S rRNA genes (Wang and Wang, 1996), and these may inflate the number of operational taxonomic units (OTUs). In this study, chimeric sequences were identified and removed using perseus, a novel program that removes chimeras from raw 454 16S rRNA sequence data, based on sequence abundances associated with pyrosequencing data (Quince et al., 2011). The uncorrected pairwise distance between aligned DNA sequences was calculated. The aligned DNA sequences were then clustered and assigned into OTUs based on a genetic distance cut off of 0.03, using the furthest neighbor algorithm as the default option. Thus a cluster was defined by sequences that differed by less than 3%. OTUs containing one or two reads are usually a result of pyrosequencing errors and these were removed from the dataset, together with OTUs detected in the negative controls. Mothur detected OTU identities up to the genus level; however an attempt was made to

identify specific species in the clustered OTUs, by performing an online BLAST search for the candidate sequences. All matching species with ≥98% identity were recorded.

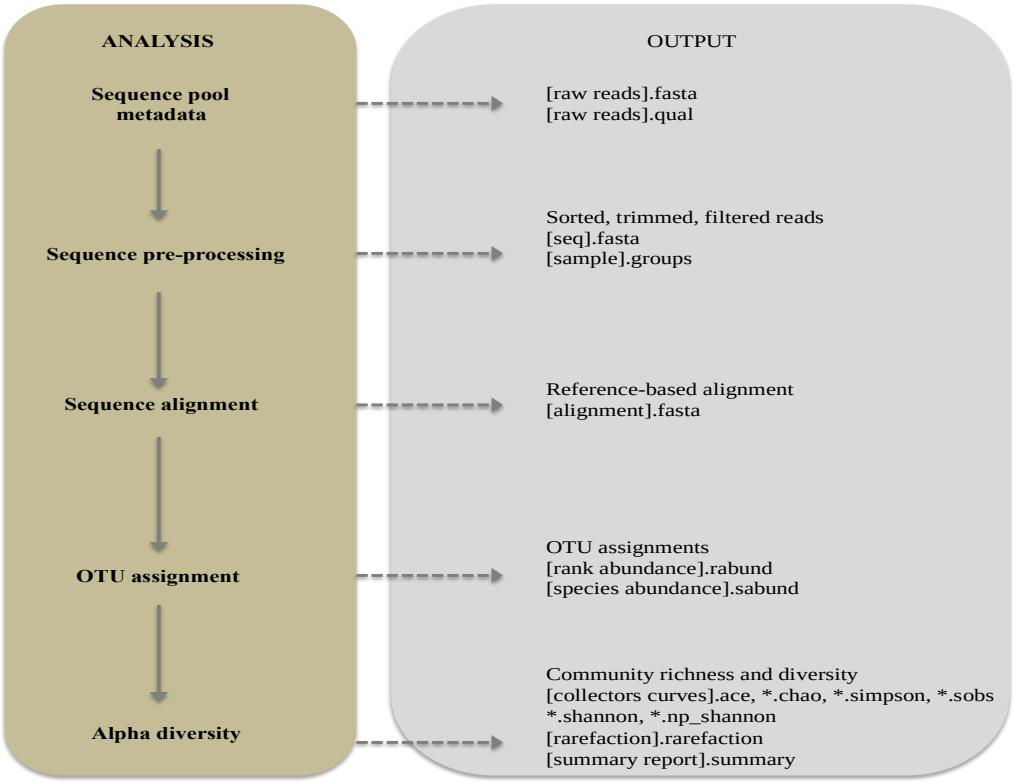


Figure 2.30 An outline of data analysis in Mothur

In this study, only alpha diversity was determined because samples were collected from the same nasopharyngeal environment (Appendix 5).

2.10.8. Phylogenetic analysis

The subsampled 16S sequences were aligned in MUSCLE (Edgar, 2004).To determine the phylogenetic relationships, a tree of the aligned sequences was generated using RAxML (Stamatakis et al., 2005) by performing 100 bootstraps.

Chapter 3: Pneumococcal carriage dynamics (cross-sectional)

3.0 Introduction

Pneumococcal acquisition is thought to play a role in the development of invasive pneumococcal disease (Chapter 1, Section 1.2). Therefore by investigating carriage we can begin to understand how disease develops and thus develop interventions to prevent it. Pneumococcal carriage is also a marker for estimating vaccine coverage and efficacy post-vaccination (Goldblatt et al., 2013). Nasopharyngeal carriage of *S. pneumoniae* is common whereas disease is rare (Brueggemann et al., 2003) and therefore studying impact of vaccines on carriage in a given population can potentially be used as an end point for licensure and implementation of pneumococcal vaccines (Goldblatt et al., 2013). The study described in this chapter was conducted in Malawi, a resource poor setting with a high burden of HIV infection and invasive pneumococcal disease (IPD). Samples were collected from Malawian children and adults using nasopharyngeal swabs (Chapter 2, sections 2.0-2.2). Previous studies in this setting, have described those serotypes associated with invasive pneumococcal disease in clinical settings (Cornick et al., 2011, Gordon et al., 2002), however very little has been reported with respect to understanding the profile of pneumococcal serotypes circulating in carriage across Malawi, with the exception of a 1997 study (Feikin et al., 2003). This study reported an 87% pneumococcal carriage rate, of which 43% were PCV7 serotypes. However, the serotypes were determined using the Quellung reaction, where only a single colony was picked from each plate. This resulted in a clear bias towards detection of those serotypes in highest abundance and would have missed cases of multiple carriage. Although multiple carriage has been reported elsewhere (Turner et al., 2013, Adetifa et al., 2012, Brugger et al., 2010), there is only one report on multiple carriage in

adults in Malawi (Glennie et al., 2012a) and none on children. In addition Malawi does not have adequate information on the profile of carriage pneumococcal serotypes circulating within the population. This chapter details the findings on serotype distribution and the prevalence of multiple carriage in both Malawian children and adults within the context of HIV infection. It is well known that HIV infection increases susceptibility to invasive pneumococcal disease through immunosuppression and impairment of the mucosal epithelial barrier integrity (Gilks et al., 1996, Nazli et al., 2010, Glennie et al., 2012a), however, the impact of HIV infection on carriage remains uncertain (Chapter 1, Section 1.4.2).

3.1. Methods

Serotyping of *S. pneumoniae* was undertaken by microarray as described in the methods section (Chapter 2, Section 2.6), by analysing the DNA extracted from a complete plate sweep of all colonies as opposed to a single colony.

3.2. Results

3.2.1. Pneumococcal carriage in children

A total of 116 nasopharyngeal swabs from 116 children (Table 3.1) in Blantyre and Karonga (Chapter 2, Section 2.1) were used for a first description of multiple carriage of pneumococcal serotypes amongst Malawian children. The samples were collected prior to the introduction of PCV13 to children in Malawi in 2011, providing important background information for estimating the impact of PCV13 post-vaccination.

Table 3.1 Characteristics of children from Karonga and Blantyre

	HIV Status				Total		Median age (yrs) (p<0.5)	
	Karonga		Blantyre		(Karonga & Blantyre)			
	Neg	Pos	Neg	Pos	Pos	Neg	Age	IQR
Male	8	0	38	22	22	46	2.3	1.0 - 4.6
Female	5	1	21	21	22	26	2.4	1.1 - 5.9
Median age (yrs) (p<0.001)					4.9	1.4		
IQR					2.7 - 8.1	0.8 - 3.1		

This table summarises HIV status by gender and sample origin and also presents median age, and interquartile range (IQR). The age range for the whole group is 0.008-23.3 years. Nasopharyngeal swabs were collected from male (n=68) and female (n=48) children, who were either HIV negative (n=72) or HIV positive (n=44).

The analysis of pneumococcal carriage in Malawian children by microarray identified 176 pneumococcal strains from 116 swabs, clearly demonstrating multiple carriage does indeed occur in Malawi. The strains were comprised of 43 distinct serotypes and nontypeables (Figure 3.1), and represent a much broader serotype distribution than previously reported in Malawi children (Feikin et al., 2003) and elsewhere (Adetifa et al., 2012). The serotypes were categorised into those covered by the vaccine (VT), not covered by the vaccine (NVT) and NTs. The NVT were further sub-categorised as high-invasive potential and low-invasive potential based on frequency stratification definitions from previous reports on invasive disease (Johnson et al., 2010). These categories were made to provide a simple means of distinguishing the 2 populations and highlight the possible NVT that could emerge

following the PCV13 impact on the pneumococcal population. All serotypes covered by PCV13 were detected and cumulatively, they represented 60% of the overall frequency of isolation, taking into account NT strains. The PCV13 coverage was estimated to be 61% when NT strains were excluded. This coverage is lower than the reported 78.3% PCV13 coverage in under-five children from the same setting (Everett et al., 2012) using invasive serotypes, which reflects the broader range of serotypes found across carriage. The NVT (39%) and NT serotypes (1%) accounted for the remaining proportion of detected serotypes. Serotype 1 was detected only three times amongst the children's data set, representing a 1.7% (3/176) frequency of isolation. In Malawi, serotype 1 is responsible for over 25% of all invasive pneumococcal disease cases (Cornick et al., 2011). Furthermore, serotypes 5 and 7F were each detected only once representing 0.6% (1/176) of the total carriage frequency. In other settings, serotypes 1, 5 and 7F have also been rarely detected in carriage although they account for the majority of invasive cases in the world (Almeida et al., 2013, Hausdorff et al., 2000). Based on the data from Malawi, the most abundant NVT serotype was 15B, followed by 13, 16F, 9L, 10B, 35B, 21, 11D, 34, 15A, and 35F (Table 3.2). Elsewhere, 9L, 15B, 16F, 21, and 35F were among the most frequently isolated NVT serotypes (Abdullahi et al., 2012b, Ercibengoa et al., 2012, Hill et al., 2008b), which is similar to our findings.

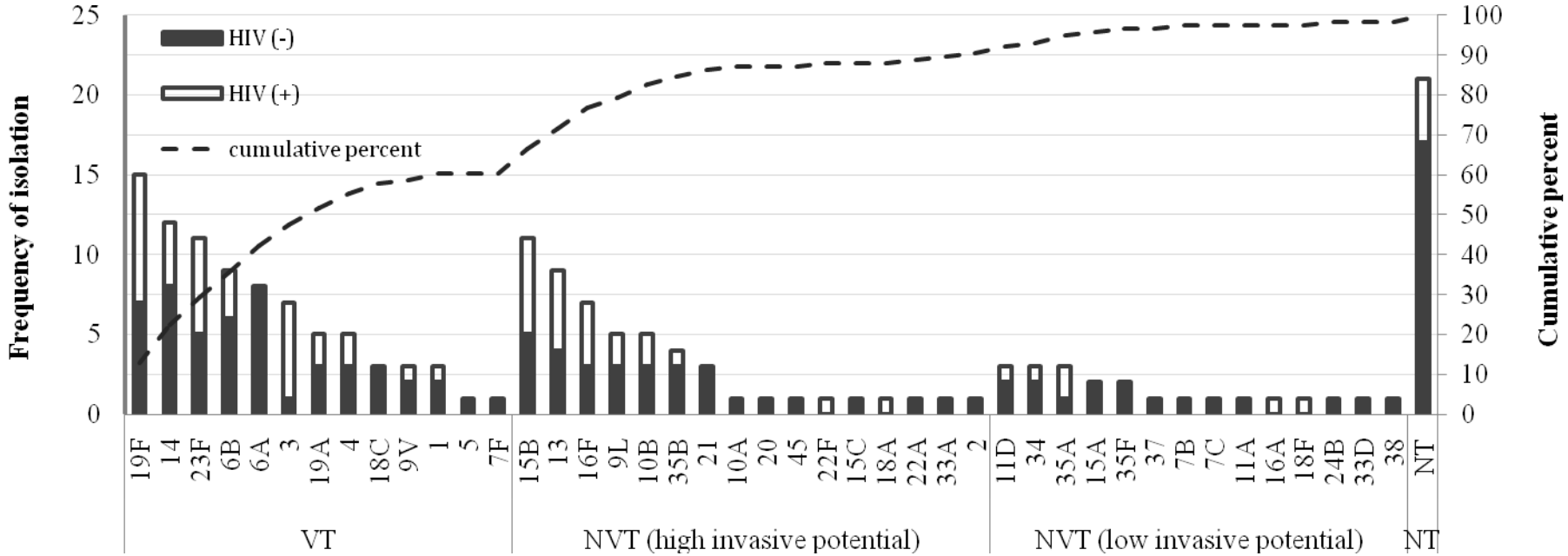


Figure 3.1 Serotype specific carriage in children

Microarray serotyping was used to determine serotypes within each sample (n=116). Results showed 43 serotypes and NTs (Tong et al.). Conventionally, NTs are grouped together since there is no specific NT classification at the moment. Cumulatively, all serotypes represented 99% abundance, while the remaining 1% was covered by NTs. A 60% PCV13 coverage was estimated from the cumulative frequency.

The data presented here demonstrate the breadth of circulating pneumococcal serotypes in Malawian children, and presents a higher frequency of serotype isolation from the sample population when compared to other settings (Abdullahi et al., 2012b, Adetifa et al., 2012). These data demonstrate that the proportion and diversity of non-vaccine serotypes circulating in Malawian children is high. Post-vaccination, these non-vaccine serotypes (Table 3.2) would be the most likely candidates for serotype replacement across carriage, and ultimately invasive disease, since non vaccine serotypes such as 15B have been reported to cause invasive diseases elsewhere (Johnson et al., 2010).

Table 3.2 Carriage of NVT serotypes in children

Serotype	**15B**	**13**	**16F**	**9L**	**10B**	**35B**	**21**	**11D**	**34**	**35A**	**15A**	**35F**
Frequency	11	9	7	5	5	4	3	3	3	3	2	2

The frequency represents the number of times each serotype was isolated out of 176.

3.2.2. Pneumococcal carriage by HIV status in children

Malawi has a high prevalence of HIV infection (UNAIDS, 2012b), and current estimates suggest 10.6 % of the population (Chapter 1, Section 1.4.1) is infected. As described previously, HIV infection predisposes patients to invasive pneumococcal disease (IPD) (Nunes et al., 2011b, Gordon et al., 2002, Carrol et al., 2007). The study reported here investigated whether HIV infection had any impact on pneumococcal carriage. One hundred and fourteen (114) pneumococcal strains were isolated from HIV negative children (n=72) and 62 from HIV positive children (n=44) (Figure 3.1). This study was not able to establish any statistically significant

difference in the total number of pneumococcal strains isolated by HIV status ($p<0.6382$). At serotype level, 39 distinct serotypes were identified within the HIV negative children whereas 22 were identified amongst HIV positive children. Twenty-one (21) serotypes were only isolated from HIV negative children and these included vaccine types 5, 6A, 7F and 18C. HIV positive children carried 4 non-vaccine types16A, 18A, 18F and 22F which were not detected in HIV negative children, and each of these serotypes was only isolated once from the entire dataset (Figure 3.1). These results show that HIV negative children carried a broader range of serotypes than HIV positive children, however, this small number of observations was considered inadequate to assess the real impact of HIV on serotype distribution. Analysis of serotype specific carriage by logistic regression showed that serotype 3 was more likely to be detected in HIV-positive (6/44) than HIV-negative (1/72) subjects ($p<0.028$; 95% CI 1.3-96.6). However, the isolation of serotype 3 from 1 HIV negative child only, raises the possibility that this may have occurred by chance, which would warrant further investigation.

Next, we investigated whether HIV infection also promoted carriage of vaccine serotypes among those infected. Vaccine serotypes are those included in the current pneumococcal vaccines; hence the need to particularly focus on their distribution as this might have an impact on disease and treatment. HIV infection has been reported to play a role in the depletion of CD4 T cells which is essential for clearance of pneumococcal strains from the nasopharynx as demonstrated in murine models of pneumococcal colonisation (Basset et al., 2007). To achieve this, carriage of three serotype categories was investigated: (i) VT, (ii) NVT or (iii) both VT/NVT (Table 3.3). Each of the three categories took into account both single and multiple carriage

events. The percentage of carriers for the VT, NVT and VT/NVT categories was 36.2% (42/116), 39.7% (46/116) and 24.1% (28/116) respectively. HIV negative subjects represented 62.1 % (72/116), and HIV positive subjects 37.9 % (44/116) of the total number of samples (Table 3.3). Although the proportion of HIV infected children in this study was higher than current HIV prevalence estimates for Malawi (Chapter 1, Section 1.4.1), more HIV positive subjects were conveniently sampled to increase the numbers for better comparison of findings with HIV negative children. Analysis by Pearson's chi2(2) t-test showed that HIV infection did not have any significant impact on carriage of VT ($p<0.1052$) and NVT ($p<0.8291$) alone. However, simultaneous carriage of both VT and NVT was slightly significant ($p<0.0388$) with HIV negative children registering higher carriage rates compared to HIV positive children. These findings highlight the fact that the relationship between carriage and HIV infection is complex and therefore currently not fully understood.

Table 3.3 Pneumococcal carriage by HIV status in children

	Carriers of:			Total (%)	Median age (yrs)* (IQR)
	VT (%)	NVT (%)	VT/NVT (%)		
HIV (-Ve)	22 (52.4	28 (60.9)	22 (78.6)	72 (62.1)	1.4 (0.8 – 3.1)
HIV (+Ve)	20 (47.6)	18 (39.1)	6 (21.4)	44 (37.93)	4.9 (2.7 – 8.2)
Total	42 (36.2)	46 (39.7)	28 (24.1)	116	***p<0.0001**

This table summarises carriage by HIV status and also presents median age, interquartile range (IQR). The age range for the whole group is 0.008-23.3 years Carriage by HIV status was investigated using three categories of pneumococcal strains: VT only, NVT only or both VT and NVT (valid for multiple serotype carriers only).

3.2.3. Pneumococcal carriage by age in children

Age is an important risk factor for pneumococcal carriage, with carriage decreasing with increasing age (Hussain et al., 2005, Adetifa et al., 2012). In this study, we investigated if carriage of pneumococcal serotypes decreased with age in Malawian children. This would help provide further data to estimate the impact of introducing PCV13 to children under the age of two years in the Malawi. Results showed that carriage decreased with age (Figure 3.2). Although sample numbers were low for older children, the results agreed with literature findings where carriage is reported to decrease with age (Hussain et al., 2005). Further analysis was conducted to establish

if age differences existed between carriers of VT and NVT types. There was no significant difference in ages for carriers of VT (median age 2.0 (0.9-4.4) years)) and NVTpneumococcal strains (median age 3.7 (1.1-3.6) years)) ($p<0.0956$) in our context. Vaccinating children aged two years and below in Malawi leaves out a larger population of older children who would act as reservoir for the spread of vaccine serotypes within the population post-vaccination, which may have a negative impact on the benefits of vaccination. However, other studies have reported that vaccinating the most affected age group has led to a significant decrease in carriage and invasive disease by vaccine serotypes in the unvaccinated population (Hammitt et al., 2006, Loughlin et al., 2014, Nzenze et al., 2013).

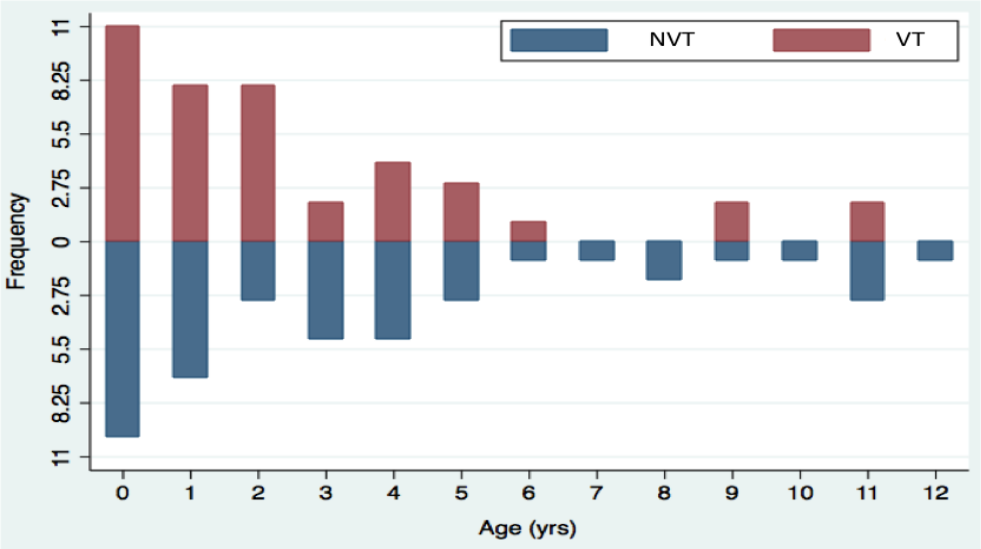


Figure 3.2 Carriage of VT and NVT serotypes by age in children

Forty-two (42) and forty-six (46) children carried the VT and NVT serotypes respectively (Table 3.3). Carriage involved either single or as multiple serotypes.

3.3. Multiple carriage in children

As previously described (Chapter 1, Section 1.2.1), multiple carriage is defined as the presence of two or more serotypes simultaneously occupying the nasopharynx. The availability of multiple carriage data provides important information on circulating serotypes in Malawi, which immediately benefit vaccine related surveillance studies. In the Malawian children's dataset, multiple carriage represented 40% (46/116) of all samples analysed, with 27% (31/116) containing two serotypes, 11% (13/116) containing three and 2% (2/116) containing four capsular types (Figure 3.3A). Multiple carriage rates are therefore very high compared to 17.3 % reported previously in Portugal among unvaccinated children (Valente et al., 2012), where serotyping was also performed using microarray. The association between serotypes in multiple carriage events was investigated across sub-groups involving two (Figure 3.3B), three (Figure 3.3C) and four serotypes (Figure 3.3D). Results from this initial analysis showed that carriage occurred randomly (Appendix 6), such that the association among serotypes in multiple carriage could not be established, however this would require further investigations to be conclusive. Using relative abundance estimates from the array data, we determined the most abundant serotypes in multiple carriage involving vaccine and non-vaccine serotypes, as these would most likely be detected by conventional serotyping techniques such as Quellung and latex agglutination. Results showed that VT and NVT serotypes were most abundant in 46.4% (13/28) and 53.6% (15/28) of multiple carriage events respectively. Thus using conventional methods, VTs would be undetected in 53.6% of these samples, thereby underestimating vaccine coverage.

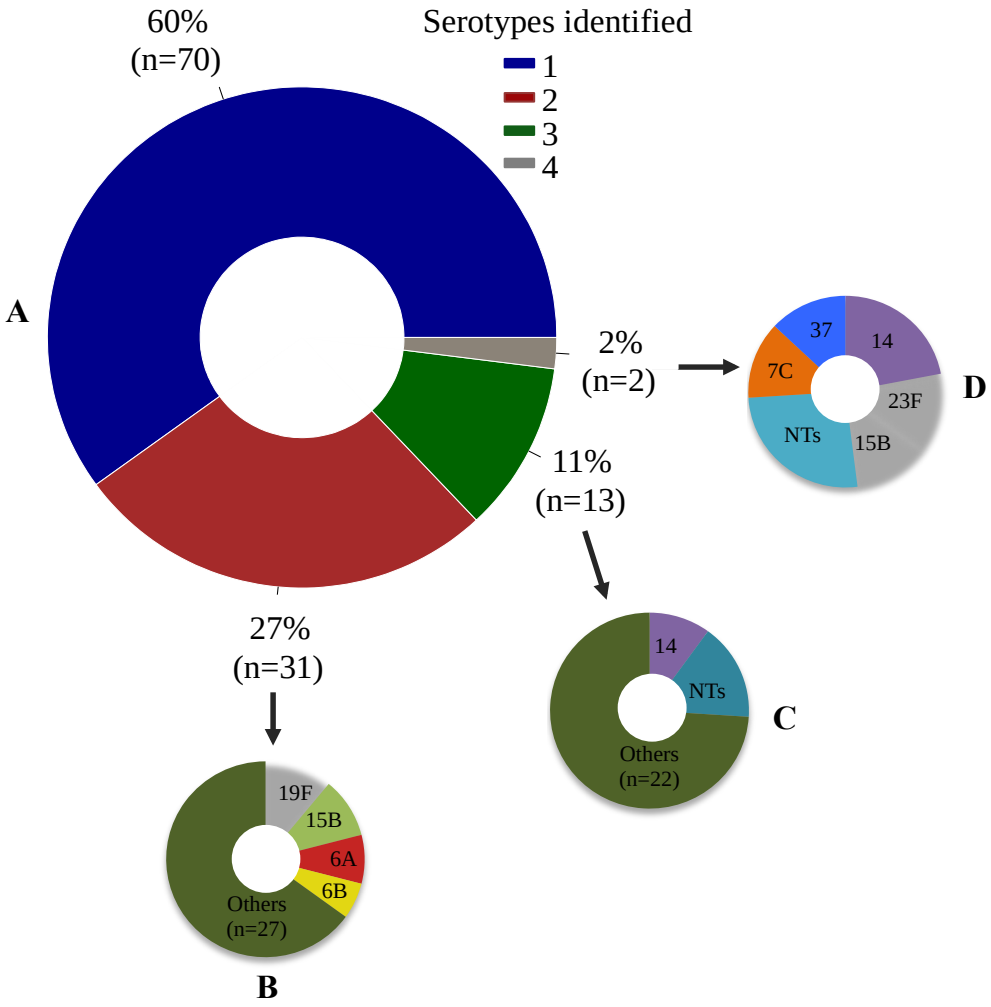


Figure 3.3 Multiple carriage in children

Carriage of multiple pneumococcal serotypes in both HIV positive and HIV negative children is shown in (A). Multiple carriage was found in 40% of samples tested, with co-colonising samples expressing two (27%) three (11%) or four (2%) capsular types. Figures B, C, and D show the specific serotype distribution in samples with two, three and four co-colonising serotypes respectively.

3.3.1. Multiple carriage by gender and age in children

In this study, the gender distribution was slightly skewed toward males (n=68) than females (n=48) (Table 3.1). Gender has previously been shown to have an impact on carriage of *S. pneumoniae* (Coles et al., 2001, Simell et al., 2008, Mackenzie et al., 2010). In this study, we investigated if gender had an impact on multiple carriage (Figure 3.4). Multiple carriage was higher in females 45.8% (22/48) than males 35.2% (24/68), however this did not reach statistical significance ($p<0.25$) (Figure 3.4). The mean number of serotypes carried by males and females was 1.5 ±0.09 and 1.6 ± 0.11 respectively ($p=0.24$), which further showed the lack of association between gender and multiple carriage in Malawian children.

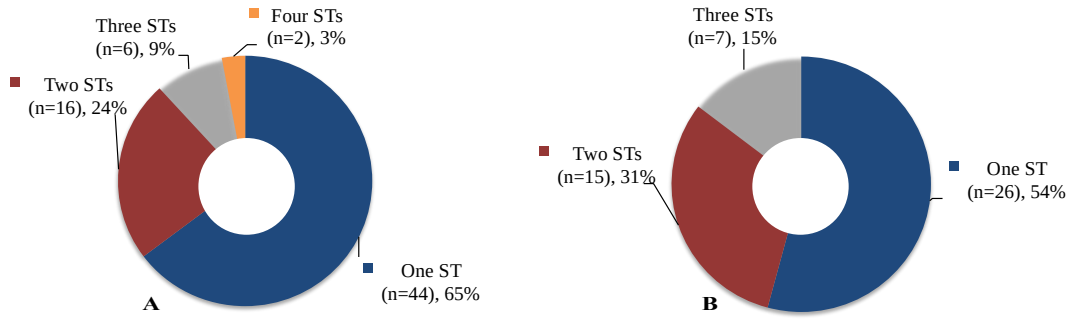


Figure 3.4 Multiple carriage by gender in children

Male and female children are represented by chart A and B respectively.

The relationship between age and carriage of *S. pneumoniae* is well documented (Hussain et al., 2005), however, it is not clear if age has a similar impact in the context of multiple carriage in children. In this study, the association between age and multiple carriage was investigated by comparing the age differences amongst single and multiple serotype carriers (Figure 3.5). The median age for single carriers (n=70) or multiple carriers (n=46) of pneumococcal serotypes was 2.6 (1.1-5.0) years

and 1.96 (1.1-3.9) years respectively (p<0.28). These results could not establish any association between age and carriage of multiple serotypes in Malawian children, although the inadequate number of samples analysed could confound this observation. However, in contrast, this thesis has shown that multiple carriage is higher in Malawian children (Chapter 3, section 3.3) compared to adults (Chapter 3, section 3.5), which highlights the fact that multiple carriage decreased with age.

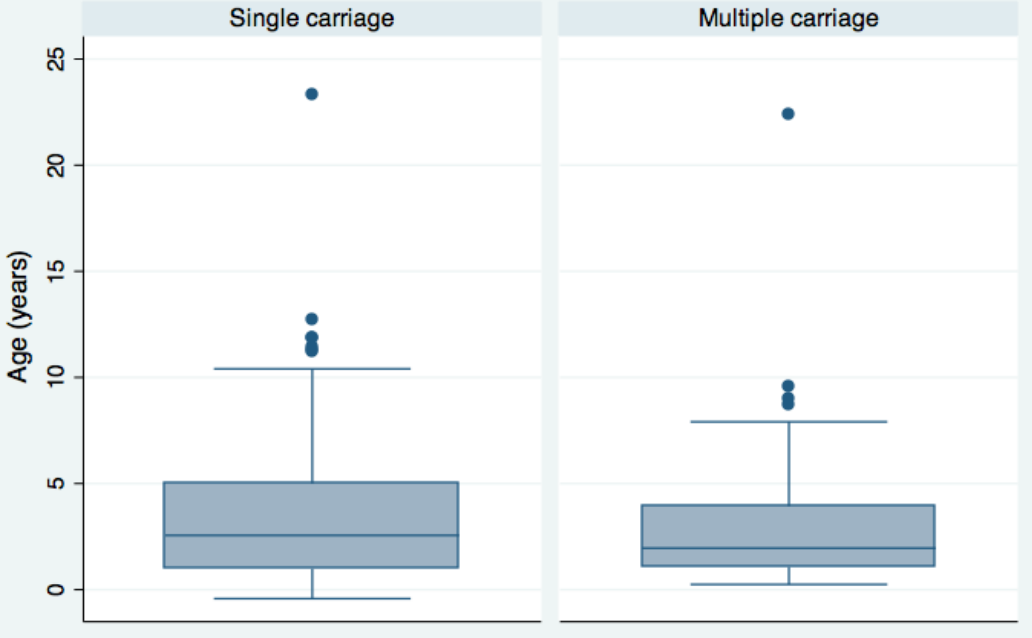


Figure 3.5 Multiple serotype carriage by age in children

3.3.2. Multiple carriage by HIV status and CD4 count in children

HIV infection impairs the function and integrity of the mucosal membrane (Nazli et al., 2010), including the nasopharynx, which compromises the ability of the membrane to clear colonising microorganisms (Glennie et al., 2010). We hypothesised that HIV infection would predispose those affected to higher prevalence of multiple carriage. To test this hypothesis, the prevalence of multiple

carriage between HIV negative and HIV positive subjects was compared (Figure 3.6).

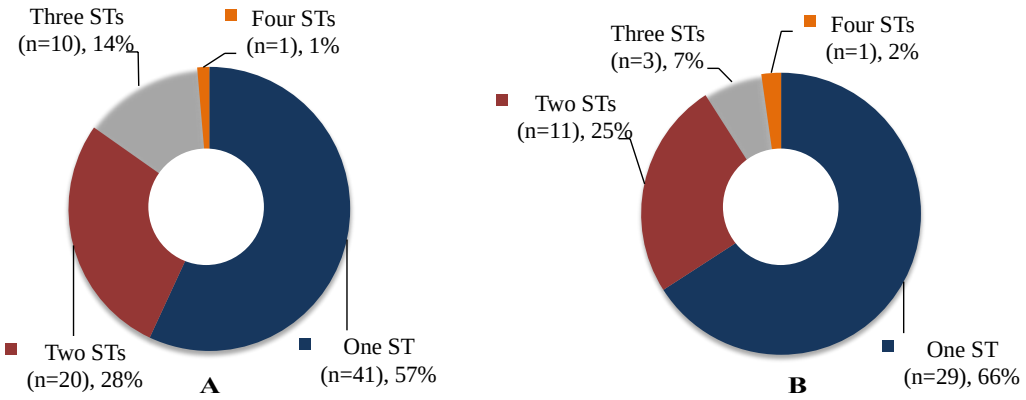


Figure 3.6 Multiple carriage by HIV status in children

Multiple carriage was 43% in HIV negative (A, n=72) and 34% in HIV positive children (B, n=44)

The majority of children in both the HIV negative (n=72) and HIV positive (n=44) groups carried one serotype followed by two, three and four serotypes. There was no difference in the mean number of serotypes carried by HIV negative (1.6) and HIV positive (1.5) children ($p<0.1659$). As shown in Figure 3.6, multiple carriage was higher in HIV negative (43%, 31/72) than HIV positive subjects (34%, 15/44), however the difference was not statistically significant ($p<0.34$), which would suggest that HIV infection did not have an impact on multiple carriage. Although a similar finding has also been reported in adults from the same setting (Glennie et al., 2012a), there was significant age difference ($p<0.0001$) in years between HIV positive children (4.9 years (2.7 – 8.2)) and HIV negative children (1.4 years (0.8 – 3.1)) which may have affected the outcome.

HIV infection lowers the peripheral blood CD4 T cell counts of the infected individuals which is associated with an increase in the incidence of bacterial pneumonia (Hirschtick et al., 1995) and pneumococcal colonisation (Glennie et al., 2012a). In this study, we investigated whether lowering CD4 count and HIV progression increased the prevalence of multiple carriage in infected children. Data on CD4 cell percentage was only reported for 79.5% (35/44) of the HIV positive children's data set. In young children, CD4 cell percentage rather than count is widely used to determine immune status, disease progression and when to start treatment (Moore et al., 2006, Gebo et al., 2004, Hulgan et al., 2007). In this subset, there were more carriers of single serotypes (n=28) than multiple serotypes (n=20) (Figure 3.7). The frequency of single serotype carriers was shown to decrease with increase in CD4 percentage; however this could be attributed to a decrease in the number of children with high CD4 % values. Unlike single carriage, there was no change observed in the prevalence of multiple carriage with CD4 percentage in the children understudy (Figure 3.7). This finding corresponds to a recent report in Malawian adults, where multiple carriage rates remained unchanged regardless of CD4 count value (Glennie et al., 2012a).

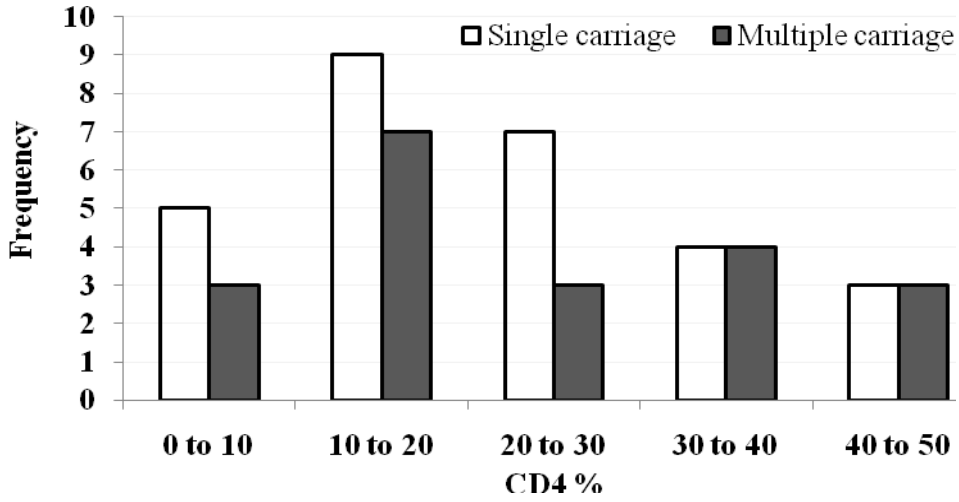


Figure 3.7 Multiple serotype carriage by CD4 percentage

CD4 cell percent values were reported for a subset (n=48) of the children's' data set.

3.3.3. Pneumococcal carriage by ART in HIV infected children

The impact of antiretroviral therapy on pneumococcal carriage dynamics in HIV positive subjects was investigated. The hypothesis was that following the possible occurrence of ART mediated immune reconstitution; the pneumococcal-specific T-cell proliferation should regulate pneumococcal carriage and colonisation by multiple serotypes. Longitudinal samples were collected from a subset of children (n=16) in Blantyre, pre- and 12 months post-ART. Results (Figure 3.8) showed that 17 pneumococcal serotypes were detected pre-ART, which was less than the 21 serotypes detected post-ART, however this difference was not statistically significant ($p<0.3368$). A significant overlap in the profile of serotypes detected pre- and post-ART was also observed. Using Pearson's chi2(2) t-test, we observed no difference in the distribution of VT or NVT in carriage between samples analysed pre- and post-ART ($p<0.298$). These findings were not able to establish the real impact of ART on

pneumococcal carriage dynamics within 12 months of use. It is likely that ART did not achieve immune reconstitution in this short period of time, which could contribute to the observed lack of impact on pneumococcal carriage. This however, requires further investigation because poor restoration of pneumococcal specific T-cell proliferative response and elevated levels of pneumococcal carriage following ART have been reported elsewhere (Cassol et al., 2013, Glennie et al., 2012a).

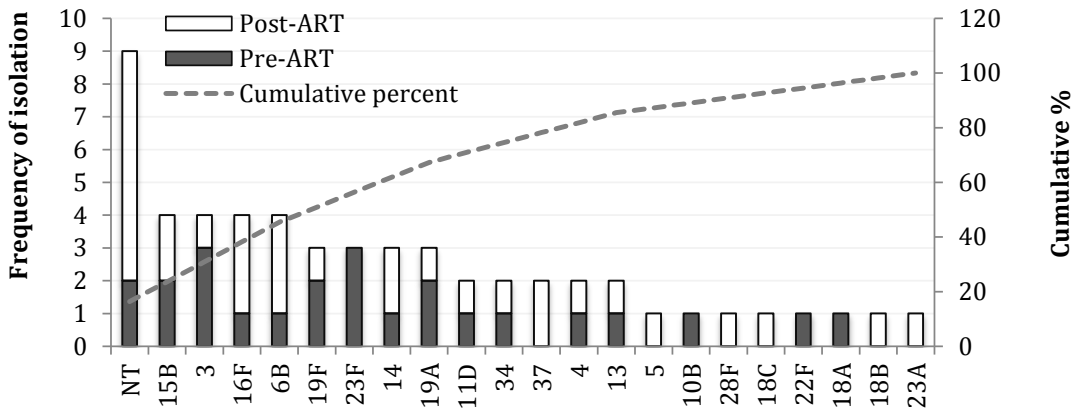


Figure 3.8 Pneumococcal carriage by ART usage in children

The children's longitudinal samples (n=16) were collected Pre- and Post-ART. An investigation into the impact of anti-retroviral therapy on multiple carriage showed that multiple carriage was significantly higher ($p<0.0325$) in samples collected 12 months Post-ART 62.5% (10/16) than those collected Pre-ART 25% (4/16) (Figure 3.9). Although the results show that high levels of colonisation with multiple serotypes were sustained following ART regimen, the sample numbers were small to be conclusive. Hence, the need for further studies to establish if indeed HIV infected children on ART are more likely to be carriers of multiple serotypes, which would make them a reservoir for person-to-person spread of *S. pneumoniae*.

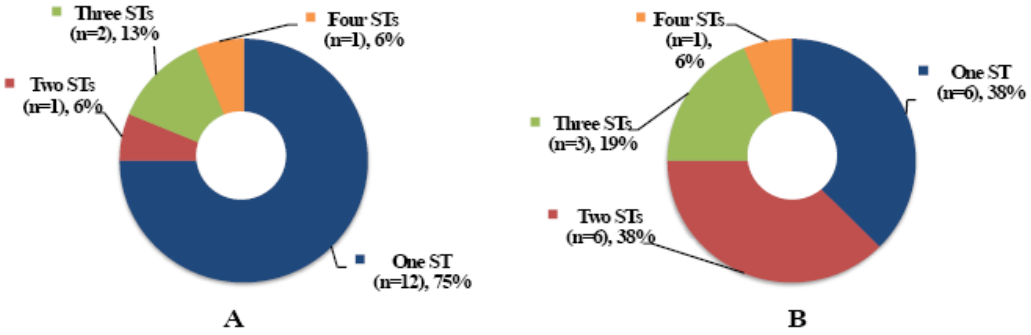


Figure 3.9 Multiple carriage by ART usage in children

Multiple carriage was 25% in samples collected pre- (A) and 62% post-ART (B) from children's longitudinal data set.

3.4 Pneumococcal carriage in adults

The adult's pneumococcal carriage data presented in this section was derived from nasopharyngeal swabs taken from a cross-sectional dataset (n=117) comprising of HIV negative (n=26) and HIV positive (n=91) adults. All the subjects presented here formed part of a slightly larger dataset (n=136) in a study reported previously (Glennie et al., 2012a). From 117 swabs, 144 pneumococcal strains were detected. Within the 144 strains, 40 distinct serotypes and NTs (Figure 3.10) were detected. Only 12 of the 13 PCV13 serotypes were identified in adults, with serotype 9V undetected. Failure to detect carriage of serotype 9V in adults may be due to chance, although serotype 9V alongside 6B, 14, 19F and 23F are recognised as paediatric serotypes (Feikin et al., 2005). Cumulatively, PCV13 serotypes represented 30% of the overall frequency of isolation taking into account NTs. Excluding NTs, PCV coverage was estimated to 38.5%, which is less than 62% reported elsewhere (Adetifa et al., 2012). This would suggest higher diversity of circulating pneumococcal serotypes in Malawian adults, however, a less sensitive latex

agglutination method and not microarray was used for serotyping in the Nigerian study, which could contribute to the observed differences. The remaining pneumococcal strains in adults (70%) comprised non-vaccine (NVT) serotypes (50%) and NTs (20%). NVT serotypes, defined as either having a high invasive potential or low invasive potential (Section 3.2.1), represented 30% and 20% of the total isolation frequency of pneumococcal serotypes respectively (Figure 3.10).

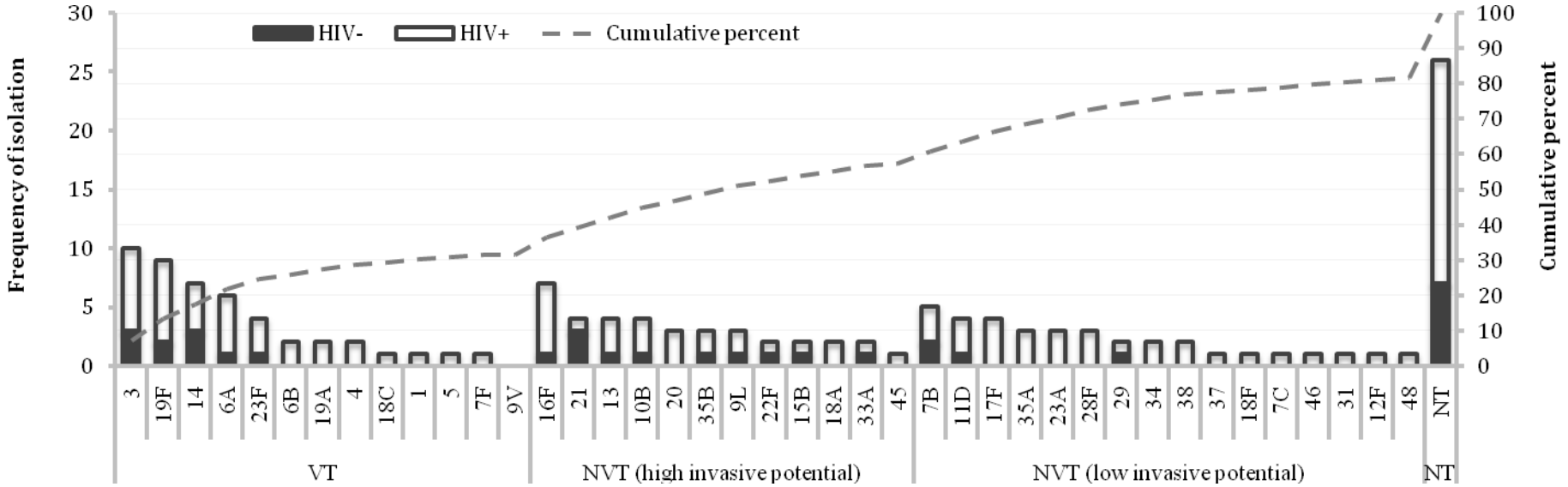


Figure 3.10 Serotype specific carriage of *S. pneumoniae* in adults

The samples (n=117) were sub-divided into HIV negative (HIV-) and HIV positive (HIV+). Serotypes present in each sample were characterised by microarray. Results showed 40 distinct serotypes (82%) and NTs (18%) representing a total frequency of isolation of 144.

Considering PCV13 serotypes alone, the most abundant serotypes ranked in order of decreasing population frequency of isolation were: 3, 19F, 14, 6A, 23F, 6B, 19A, 4, 18C, 1, 5 and 7F.Serotype 1, 5 and 7F were detected once in single serotype adult carriers in the entire data set, each representing 0.7% (1/144) of the total frequency of serotype isolation. As previously described, both serotypes 1 and 5 are the most common causes of invasive disease in Malawi (Everett et al., 2012, Cornick et al., 2011) and elsewhere (Esteva et al., 2011), however they are rarely detected in carriage (Almeida et al., 2013). The NVT serotypes present in adults and their frequency of isolation are shown in the figure below (Figure 3.11). Serotypes 16F, 10B, 11D, 13, 17F and 21 were the most commonly isolated NVTs in adults representing a total frequency of isolation of 28.

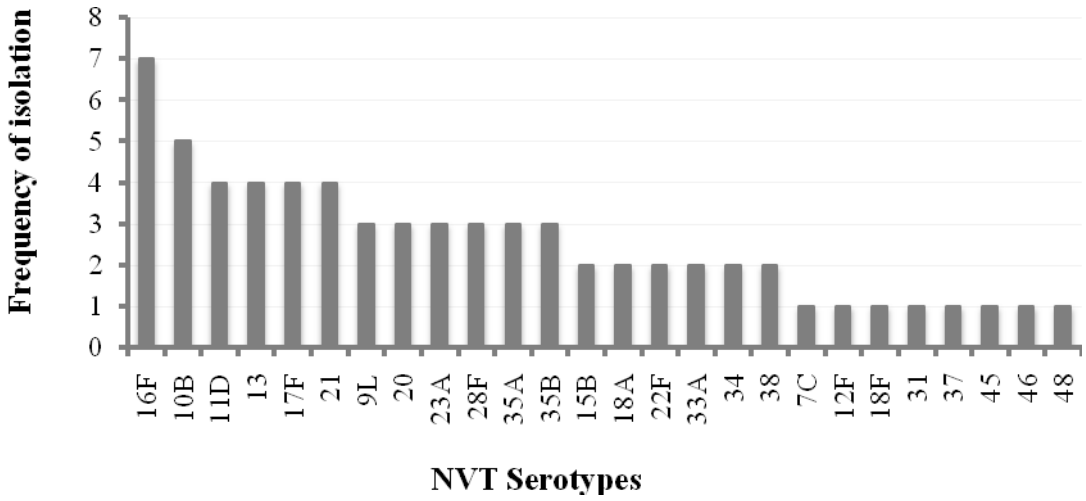


Figure 3.11 Carriage of nonvaccine serotypes in adults

3.4.1 Pneumococcal carriage by HIV status in adults

By HIV status, the adult's dataset was grouped into HIV negative (n=26) and HIV positive (n=91) individuals (Chapter 2, Section 2.2). The distribution of pneumococcal serotypes in the two cohorts was investigated to establish if HIV infection had any impact on the observed diversity of pneumococcal carriage. Thirty-two (32) pneumococcal strains were isolated from HIV negative adults (n=26) and 112 from HIV positive adults (n=91) (Figure 3.10). There was no statistical difference in serotype distribution by HIV infection ($p<1.0$), however, because of the limited sample size, this may not be conclusive. Seventeen (17) distinct serotypes were identified within the HIV negative individuals whereas 40 were identified amongst HIV positive individuals. Within those 40 serotypes, all the serotypes detected in the HIV negative cohort were also present which demonstrated an overlap across carriage by HIV status. The results for serotype distribution after sub-categorising HIV positive individuals by CD4 count are presented in Figure 3.12.Of the 91 HIV positive individuals, the CD4 count data was available for 79 subjects. Samples were stratified into two groups: CD4≤250 (n=44) and CD4>250 (n=35). Within this subset 40 pneumococcal serotypes were identified and 19 of those serotypes were common to both cohorts. Those with CD4≤250, had 13 serotypes (10B, 18A, 5, 7C, 7F, 12F, 13, 18F, 20, 22F, 31, 37, 46) which were only isolated from this cohort, whereas 8 serotypes (21, 9L, 1, 4, 6B, 18C, 33A, 45) were identified only in those with CD4>250. The large proportion (47.5% (19/40)) of serotypes common to both cohorts suggested no difference existed in the distribution of serotypes. This was further supported by calculating the mean number of serotypes carried per individual, with those having CD4≤250 registering 1.25 (1.07-1.42), similar to those with CD4>250 who registered 1.13 (1.03-1.23).

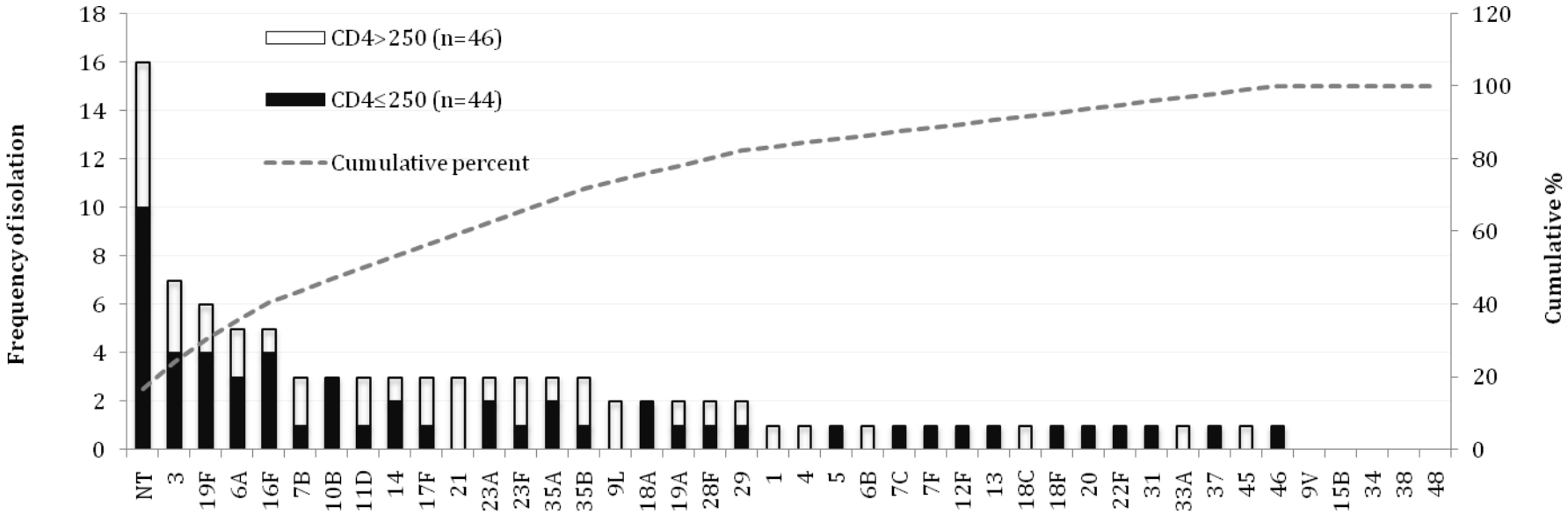


Figure 3.12 Pneumococcal carriage by CD4 count in adults

Two groups of samples from HIV positive adults were compared based on whether their CD4 count was less than or equal to 250 (CD4≤250) or greater than 250 (CD4>250). Overall, 40 pneumococcal serotypes (76%) and NTs (24%) were identified.

3.4.2 Pneumococcal carriage by ART in HIV infected adults

In this study, a broader range of pneumococcal serotypes was observed in those individuals with low CD4 counts. Previous findings have shown an increase in colonisation rates following a drop in CD4 count (Rodriguez-Barradas et al., 1997). We investigated whether ART usage had any impact on pneumococcal carriage in adults (Figure 3.13). The dataset comprised 19 adults on ART (ART+) and 35 ART naïve adults. The dataset yielded 72 pneumococcal strains with 28 distinct serotypes (67%) and NTs (33%), by microarray. From the ART negative cohort, 46 pneumococcal strains were detected, which comprised 24 distinct serotypes and NTs; while the ART positive group yielded 26 pneumococcal strains, comprising 13 distinct serotypes and NTs. Further analysis showed that 15 serotypes (14, 10B, 23A, 4, 6B, 7C, 12F, 15B, 17F, 18A, 18C, 20, 22F, 37, 38) were only detected in adults not exposed to ART, while 4 serotypes (5, 6A, 33A, 45) were only isolated from adults on ART. The data in our study showed that subjects who were ART naïve carried a broader range of pneumococcal serotypes than the ART+ group, however the difference was not statistically significant ($p<0.2541$). Similarly, there was no difference in the mean number of serotypes carried between subjects who were ART naïve (1.31, 1.13-1.50) and those exposed to ART (1.37, 1.08-1.66). Given the limitation with sample numbers, these findings reported here would warrant further investigation to establish the actual impact of ART on carriage.

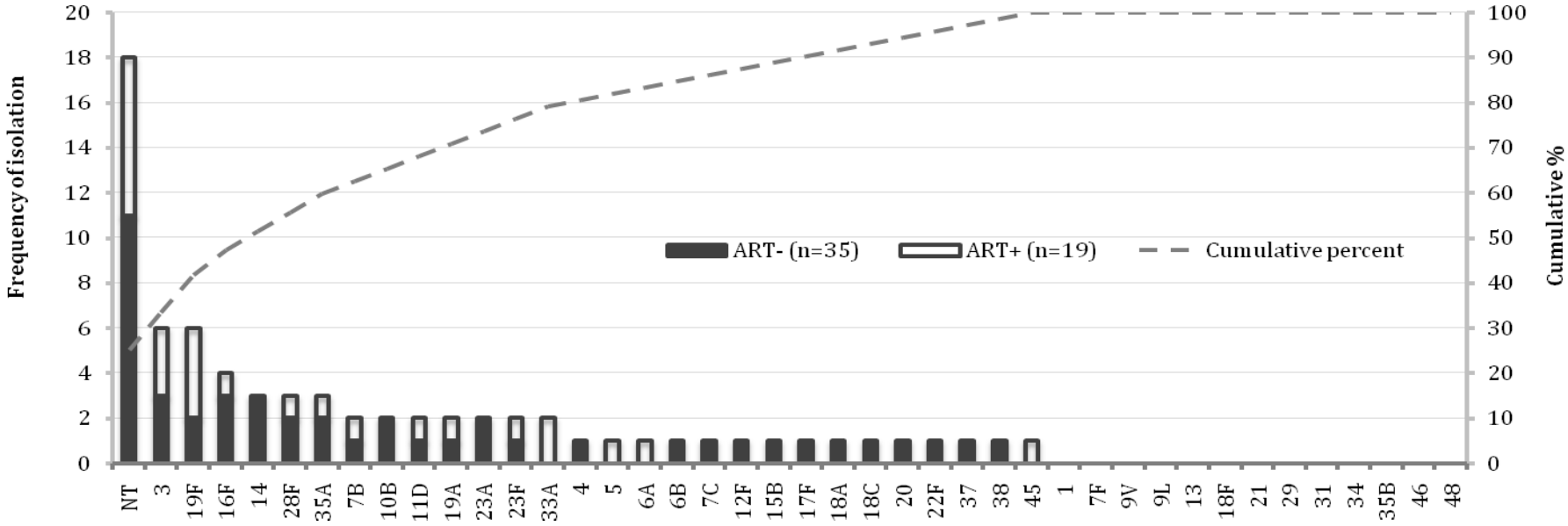


Figure 3.13 Pneumococcal carriage by ART in HIV infected adults

Two groups were compared based on whether they were on anti-retroviral therapy (ART+) or not. A total of 72 pneumococcal strains were isolated, with 28 distinct serotypes (67%) and NTs (33%) detected.

3.5 Multiple carriage in Malawian adults

Pneumococcal carriage in children is higher than in adults (Hussain et al., 2005, Adetifa et al., 2012). In this study, we investigated whether this was true with respect to multiple carriage. Results showed that 19% (22/117) of the adult population (Figure 3.14A) carried at least two serotypes in their nasopharynx. The maximum number of serotypes carried at any one time was 3 and this was observed in 4% (5/117) of the adult population. The majority of multiple carriage in adults involved 2 serotypes (15% (17/116)). Compared with children (Section 3.3), the results showed that multiple carriage was lower in adults by 21%. By HIV status, multiple carriage rates were similar in HIV negative (16%, 4/26) (Figure 3.14B) and HIV positive (20%, 18/91) (Figure 3.14C) adults, which supports previous observations where multiple carriage was not affected by HIV status (Glennie et al., 2012a). The same study reported an increase in colonisation rate with a decrease in CD4 count in adults (Glennie et al., 2012a). In this study, we investigated if CD4 count had any impact on multiple carriage in adults. To achieve this, the two sub-groups with CD4≤250 (n=44) and CD4>250 (n=35) were compared. Multiple carriage was detected in 18.2% (8/44) of the samples from the CD4≤250 group and 17.1% (6/35) in samples from individuals with CD4>250, demonstrating no difference in carriage of multiple serotypes. Similarly, multiple carriage was also unaffected by ART usage, with those exposed to ART having a 31.5% (6/19) while 28.6% (10/35) was observed in individuals who were ART naïve. This is in agreement with previous findings where elevated levels of pneumococcal colonisation were maintained in adults established on ART (Glennie et al., 2012a).

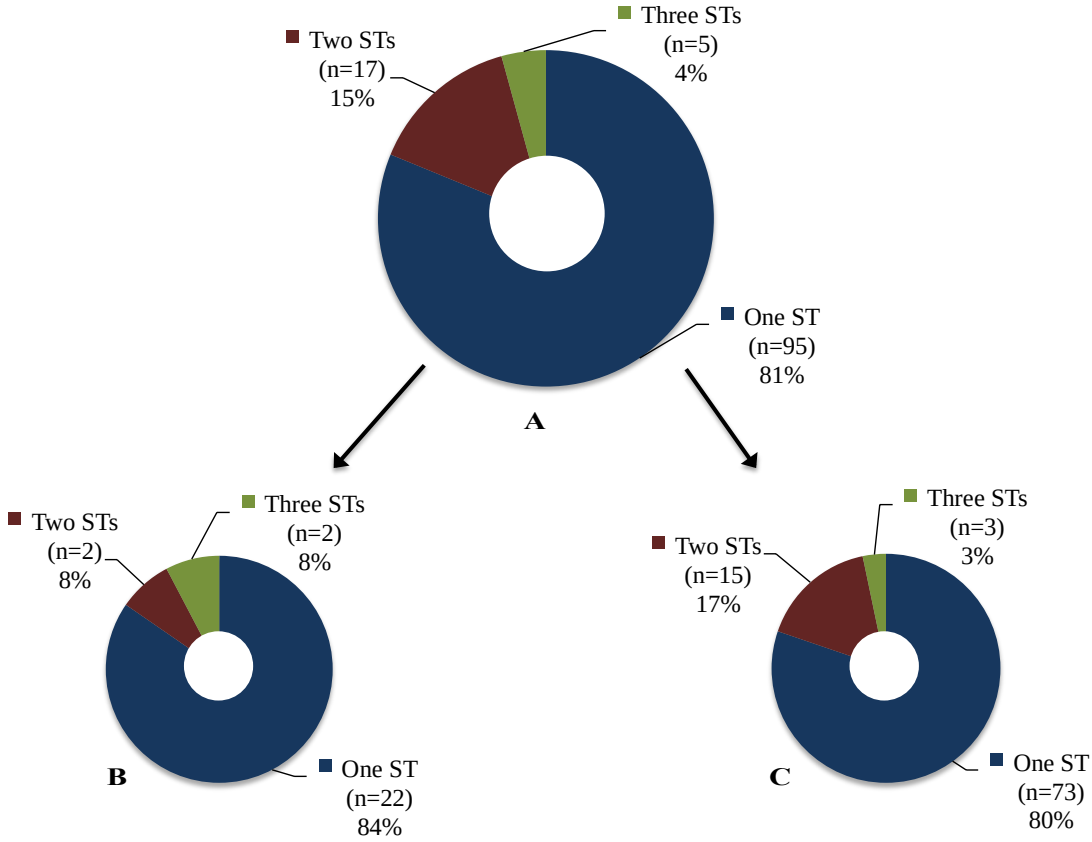


Figure 3.14 Multiple carriage in adults

The prevalence of multiple carriage in Malawian adults (A) was ~19 % (22/117). By HIV status, multiple carriage was 16% (4/26) in HIV negative subjects (B) and 20% (18/91) in HIV positive subjects (C).

Further assessment of serotype associations in multiple carriage (Figure 3.15) showed 4 serotypes (10B, 31, 29, and 23A) that were only isolated from multiple carriage events involving three serotypes, while 18 serotypes (3, 33A, 23F, 13, 21, 9L, 38, 6B, 28F, 4, 35A, 45, 22F, 7B, 37, 19A, 14, 19F) were only isolated from multiple carriage events involving two serotypes. This was done to characterise possible associations among serotypes in multiple carriage, but due to a small number of samples carrying multiple serotypes, we were not able to make such deductions, hence the need for further investigations. However, nontypeable

serotypes (Chapter 1, section 1.1) were isolated from both events of multiple carriage. Cumulatively NTs were present in 43.8% of all multiple carriage events (Figure 3.15). NTs have been widely isolated in other carriage studies (Marsh et al., 2010, Scott et al., 2012), and in this study, we found that NTs appear in multiple carriage events more than any other serotype ($p<0.0058$) in Malawian adults. By carrying high proportion of NTs, Malawian adults may contribute to the development of antibiotic resistance in Malawi because the presence of NTs in multiple carriage with encapsulated serotypes would promote the dissemination of antibiotic resistance genes (Marsh et al., 2010, Hauser et al., 2004, Chewapreecha et al., 2014).

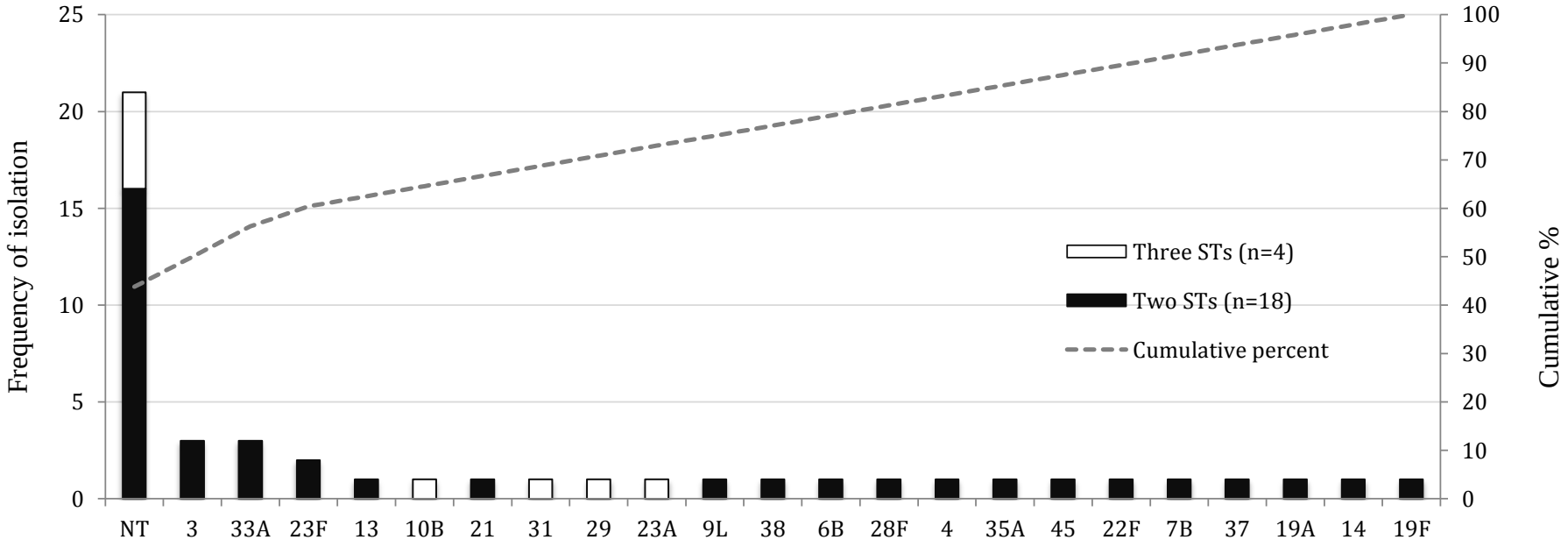


Figure 3.15 Serotypes isolated from multiple carriage in adults

Nontypeable serotypes (NT) were most likely to be detected in multiple carriage with encapsulated *S. pneumoniae* and cumulatively, NTs represented 45% of the total frequency of isolation.

3.6 Discussion

Historically, there has been a reliance on serological techniques such as the Quellung reaction and latex agglutination for pneumococcal serotyping (O'Brien and Nohynek, 2003). The limitation of using serological techniques in serotyping is that they are expensive, labour intensive, and are biased towards detecting the most abundant serotype in subjects carrying multiple serotypes. In this study, the diversity of pneumococcal carriage in Malawian children and adults was determined by microarray (Chapter 2, section 2.6), which does not only detect serotypes, but also carriage of multiple serotypes and their relative abundance (Turner et al., 2011).

A broad range of carriage pneumococcal serotypes in Malawian children and adults. Serotype distribution across children (Section 3.2.1) and adults (Section 3.4) was equally broad with a total of 43 and 40 distinct serotypes and NTs detected respectively. This diversity was higher than any previously reported study in Malawi (Feikin et al., 2003) or elsewhere in Africa (Adetifa et al., 2012). Based on carriage data, PCV13 serotypes coverage was estimated at 31% in adults, which was lower than the 60% in children, suggesting that the vaccine would likely be more effective in children compared to adults. Based on invasive disease surveillance data, PCV13 coverage in Malawi was reported to be 88% (Cornick et al., 2011). The difference in PCV13 coverage estimates between carriage and invasive strains is to be expected as there is a more diverse range of serotypes in carriage compared to those commonly associated with invasive disease. Failure to adequately account for serotype 1 in carriage could also contribute to the apparent low coverage of PCV13. Therefore coverage estimates based on carriage data should be interpreted in this context when considering using them as supplementary or alternative endpoint for licensure of

vaccine in a given population (Goldblatt et al., 2013).

The low PCV13 coverage estimates based on carriage also suggest that a large proportion of NVT serotypes would remain in circulation post-vaccination. For a resource poor setting such as Malawi, this is of concern because an increase in carriage of NVT post PCV introduction (Obaro et al., 1996), may be followed by an increase in NVT IPD (Vestrheim et al., 2010). Although our data suggest a high likelihood of serotype replacement occurring in Malawi post vaccination (Link-Gelles et al., 2013), it is not possible to predict which serotypes those might be, because the reasons a particular NVT serotype increases after vaccine introduction remain unclear (Link-Gelles et al., 2013). Previously, it has been reported that a non-vaccine serotype 22F (with low carriage prevalence) was associated with replacement disease post PCV7 (Flasche et al., 2011b). In our dataset, serotype 22F was isolated in carriage from both children and adults, making it one of the most likely serotypes to cause replacement disease post PCV13 in Malawi. The predicted post vaccination effect however can only be verified by conducting post vaccination surveillance of pneumococcal serotypes in carriage and/or disease.

Nontypeable serotypes represented <2% and 18% of all serotypes detected in Malawian children and adults respectively. In other settings, NT carriage rates of up to 18% (n=97) (Mackenzie et al., 2009, Marsh et al., 2007) and 6% (n=218) (Ercibengoa et al., 2012) have been reported in children, which were higher than Malawian children for similar sample sizes. In Thailand (n=234, mother infant pairs), carriage of NTs in children and adults (mothers) was 10.5% and 29.6% respectively (Turner et al., 2012), which was higher than Malawi, although in both settings, NT

carriage rates were higher in adults than children. The reasons for higher carriage of NTs in adults than children remain unknown. Further observation showed that NTs were frequently isolated from multiple carriage with encapsulated pneumococcal strains in our setting (Section 3.5), which may increase the chance of genetic exchange with encapsulated pneumococcal strains (Chewapreecha et al., 2014).

Multiple carriage is higher in Malawian children than adults. The prevalence of multiple carriage in Malawian children (Section 3.3) was 40% (46/116), higher than that observed elsewhere (Valente et al., 2012). The prevalence of multiple carriage in Malawian adults (Section 3.5) was 19% (22/117), which was lower than the prevalence in children. The maximum number of serotypes isolated from multiple carriage in children was 4, isolated from 1.7% (2/116) of the children's population. In adults, the maximum number of serotypes carried at any one time was 3 and this was observed in 4% (5/117) of the entire adult population, whereas in children multiple carriage involving 3 serotypes was detected in 11 % (13/116) of the children's population. The majority of multiple carriage events across children and adults involved 2 serotypes, which was also higher in children (27% (31/116)) than adults (15% (17/117)). The observed difference in multiple carriage between children and adults could be attributed to the developing immune system in children, which might not be able to regulate pneumococcal carriage as well as the developed immunity in adults (Granat et al., 2009, Glennie et al., 2012b). Other factors such as having young siblings in the home and breast-feeding, which impact on carriage in children (Regev-Yochay et al., 2004b, Reisman et al., 2013), may also play a role in the observed higher multiple carriage rates in children.

In addition to promoting horizontal gene transfer (Donkor et al., 2011), multiple carriage has also been associated with an increase in the density of pneumococcal carriage in the nasopharynx (Brugger et al., 2010). Recently, studies have shown a correlation between an increase in the density of *S. pneumoniae* in carriage and pneumonia in adults (Albrich et al., 2012, Vu et al., 2011), suggesting density could be used as a diagnostic marker. However, in a setting with high rates of multiple carriage, the diagnosis of pneumoniae based on bacterial density needs to be interpreted with caution (Shak et al., 2012). A recent study has shown a decrease in multiple carriage following vaccination (Valente et al., 2012). It remains to be see whether multiple carriage would decrease post-vaccination in Malawi. A decrease in multiple carriage would consequently decrease the available pool of genes required for genetic recombination (Chapter 5, Section 5.2.1). However, recombination rates are still likely to accelerate among vaccine serotypes under vaccine selection pressure (Brueggemann et al., 2007a, Croucher et al., 2013) as a survival mechanism. The consequence of which is to give rise to new strains that may not be responsive to vaccination. Malawi introduced routine pneumococcal vaccination with PCV13 as part of the infant expanded programme on immunisation in November 2011 (Everett et al., 2012), targeting children aged 2 years and below. The PCV13 schedule is recommended for children aged between 2-71 months (Table 3.4) and adults aged 50 years and above (http://www.cdc.gov). In Malawi, the Ministry of Health recommends administration of a 3-dose PCV13 schedule to children at 6, 10 and 14 weeks old. We have shown in our study, that carriage of PCV13 serotypes is similar in Malawian children across the age range of 0-12 years. However, vaccinating children under the age of 5 years has been reported to be effective at reducing transmission to older children and adults (Section 3.2.3), thereby ensuring herd

immunity. PCV13 is expected to reduce carriage and invasive disease by VT serotypes in the vaccinated population in Malawi, however, elimination of VT serotypes is followed by carriage of less abundant non-vaccine serotypes, which has led to NVT invasive disease in other settings (Singleton et al., 2007, Flasche et al., 2011b, Weinberger et al., 2011). Further carriage studies post-vaccination would be recommended to assess the impact of vaccine introduction.

Table 3.4 PCV13 routine vaccination schedule

Age at first dose (months)	Primary PCV13 series	PCV13 booster dose
2-6	3 doses	1 dose at age 12-15 months
7-11	2 doses	1 dose at age 12-15 months
12-23	2 doses	-
24-59 (healthy)	1 dose	-
24-71 (children with immunocompromising conditions)	2 doses	-

A recommendation by the advisory committee on immunization practices (ACIP), United States, 2010 (*http://www.cdc.gov*).

Impact of HIV infection on pneumococcal carriage and multiple carriage in Malawian children and adults. We showed that HIV negative children (Section 3.2.2) carried a broader range of serotypes compared to HIV positive children. Of all the serotypes detected in children's data set, only serotype 3 was more likely to be carried in HIV infected than un-infected individuals. Although serotype 3 was

isolated in adults, its association with HIV infection was not observed (Section 3.4.1). Such a finding therefore requires further investigations; since there is a possibility it may have occurred by chance. HIV infection did not have a significant impact on the prevalence of multiple carriage ($p<0.34$) although it was higher in HIV negative (43%, 31/72) than HIV positive children (34%, 15/44). There was also no significant difference ($p<0.5447$) in the prevalence of multiple carriage between HIV negative (15%, 4/26) (Figure 3.14B) and HIV positive (20%, 18/91) (Figure 3.14C) adults, which suggested the lack of association between HIV infection and multiple carriage in Malawian children and adults, as previously reported in Malawian adults (Glennie et al., 2012a).

Pneumococcal carriage diversity in adults was higher in the HIV positive compared to HIV negative adults (Section 3.4.1). However, there was no significant difference in the diversity of serotypes in carriage by HIV status ($p<0.7090$) because of a huge overlap in serotypes detected in the two groups. This finding is consistent with a previous carriage study in the USA (Rodriguez-Barradas et al., 1997), where no significant difference was observed in the carriage prevalence of any serotype between HIV-infected and –uninfected adults. Among the HIV positive adults, carriers with low CD4 count (CD4≤250) had a wider range of serotypes than those with high CD4 counts (CD4>250). It is known that CD4 counts are lower in HIV positive patients and previous work in Malawi has shown colonisation rates increase with decreasing CD4 counts (Glennie et al., 2012a). HIV infection has been reported to impair the innate and adaptive immune system and also compromises the integrity and function of the epithelial barrier (Glennie et al., 2010). This renders the host mucosal barrier ineffective at regulating carriage and could explain why there is

broad range of pneumococcal serotypes carried in HIV positive subjects, particularly those with low CD4 counts.

Impact of ART on carriage and multiple carriage in HIV infected children and adults. HIV infected subjects have a depleted number of peripheral blood CD4 T cells which (Cassol et al., 2013) is associated with prolonged colonisation of mucosal pathogens such as *S. pneumoniae* (Glennie et al., 2010). Anti-retroviral therapy (ART) reconstitutes the immune system by increasing CD4 T cell levels (Le et al., 2013), which is expected to enable the host immune system mount a better response against pneumococcal carriage. The usage of ART in children (Section 3.3.3) did not result in any significant difference ($p<0.3368$) in the distribution of pneumococcal serotypes in samples collected pre- and post-ART. Similarly, ART did not have any impact on serotype distribution in adults ($p<0.9435$) although subjects who were ART naïve carried a broader range (25, n=35) of pneumococcal serotypes than the group on ART (14, n=19). Having shown that ART did not have an impact on serotype distribution in both children and adults in our context, we investigated whether this also applied to carriage of multiple serotypes. Whilst ART showed no effect on multiple carriage in the adult population (Section 3.4.2), there was a significant increase of multiple carriage in HIV positive children ($p<0.038$), following ART introduction (Section 3.3.2). Recently, it has been reported that although there is an increase in CD4 cells following ART usage, there is a persistence of defects in pneumococcal specific T cell proliferation (Glennie et al., 2012a) which results only in partial reconstitution of the immune system (Cassol et al., 2013). This could contribute to the lack of change in multiple carriage prevalence; following ART usage in Malawian adults and children, however there is

need to establish if immune reconstitution was achieved in this context before such conclusions can be made. In addition, further analysis on a larger dataset and also after prolonged usage of ART would be recommended, given a study in Brazil reported lower risk of colonisation by *S. pneumoniae* following prolonged ART usage (Nicoletti et al., 2007).

3.7 Conclusion

This chapter has demonstrated the carriage of a broad range of serotypes amongst children and adults in Malawi. These serotypes included those covered by the current conjugate vaccine (PCV13), with an estimated 60% and 30% coverage in children and adults respectively; and a wide variety of non-vaccine serotypes. The coverage is lower than 88.8% estimate based on invasive disease in Malawian children ranging from 2 months to 16 years old (Cornick et al., 2011). This difference could be attributed to prevalence of serotype 1 which was present at very low abundance in carriage in our dataset yet it is the most commonly isolated serotype from invasive disease in Malawi. This suggests that estimating coverage based on carriage data only can be misleading. Our dataset had a large proportion of non-vaccine serotypes detected in both children and adults and potentially, these non-vaccine serotypes will be freely transmitted within the population. This may create a perfect environment for serotype replacement post vaccination, which may affect the benefits of PCV13 introduced in Malawi in 2011, if it translates to disease by non-vaccine serotypes (Weinberger et al., 2011). The data has also demonstrated for the first time in Malawi, a high prevalence of multiple carriage in Malawian children (40%) and a relatively lower prevalence in adults (19%). It is therefore recommended that nationwide post-vaccination pneumococcal carriage and disease surveillance be

established to monitor the impact of vaccination. By HIV status, HIV negative children carried a broader range of serotypes than HIV positive children, whereas a significant overlap in carriage serotypes was observed in adults by HIV infection. HIV infection did not show an impact on the prevalence of multiple carriage in Malawian children and adults. There was also no impact of ART usage on multiple carriage in adults, however, multiple carriage was significantly higher in samples collected from children post-ART than pre-ART. The data on carriage by HIV status in our setting requires further investigation because to date the relationship between carriage and HIV infection is still not fully understood. A summary comparing the findings in adults and children is presented in table 3.5.

Table 3.5 Summary of carriage findings in children and adults

Variable **General**		**Children (n=116)**		**Adults (n=117)**	
1	Total isolation freq.	176		144	
2	Serotypes	43 (99%; cum. percent)		40 (82%; cum. percent)	
3	Nontypeables	1% (cum. percent)		18% (cum. percent)	
4	PCV coverage	60%		31%	
5	Multiple carriage	40%		19%	
HIV status		**HIV (Costa et al.) (n=72)**	**HIV (+ve) (n=44)**	**HIV (Costa et al.) (n=26)**	**HIV (+ve) (n=91)**
1	Total isolation freq.	114	62	32	112
2	Serotypes	39	22	17	40
3	Nontypeables	0.8%	0.2%	5%	13%
5	Multiple carriage	43%	34%	16%	20%
ART Usage			**ART (-) (n=16)**	**ART (+ve) (n=16)**	**ART (-) (n=19)**	**ART (+ve) (n=35)**
5	Multiple carriage		25%	62%	31.5%	28.6%

Chapter 4: Pneumococcal carriage dynamics (longitudinal) by microarray and Quellung serotyping

4.0. Introduction

Children are colonised by a number of different strains of *S. pneumoniae* from birth (Hussain et al., 2005, Lloyd-Evans et al., 1996). The duration of nasopharyngeal carriage can vary with each serotype. Generally in children, a single serotype can be carried for up to three months (Hill et al., 2008b, Turner et al., 2012). Factors that affect the duration of serotype carriage include the host immunity (Goldblatt et al., 2005). The duration of carriage for all serotypes has been shown to decrease with age (Cobey and Lipsitch, 2012) due to the acquisition of nonspecific immunity to pneumococcus (Cobey and Lipsitch, 2012, Hogberg et al., 2007). This immunity is thought to be largely dependent on CD4+ T helper -17 cells (Malley et al., 2005, Lu et al., 2008). Other factors such as competitive interaction with other colonising bacteria in the nasopharynx (Bosch et al., 2013), viral co-infections (Diavatopoulos et al., 2010), as well as thickness of the capsule (Hogberg et al., 2007) have all been shown to impact on carriage duration. It has also been reported that the duration of carriage is inversely proportional to the invasive capacity of a given serotype (Sleeman et al., 2006). Having analysed the cross-sectional carriage of *S. pneumoniae* in both adults and children (Chapter 3), we then investigated how carriage of *S. pneumoniae* varied with time in an individual. The aim was to understand levels of pneumococcal transmission and describe pneumococcal genetic diversity over-time given their relevance to disease control. This chapter therefore discusses carriage of *S. pneumoniae* over time in infants less than 12 months old, from a rural population in Karonga, taking advantage of a larger study of

pneumococcal carriage undertaken at the Karonga Prevention Study (Chapter 1, section 1.0).

4.0.1 Methods

The detailed methods for this section are outlined in chapter 2. Briefly, the profile of carriage pneumococcal serotypes was defined using a subset of longitudinal samples (n=16) randomly selected from a larger dataset of 158 participating infants in Karonga (Chapter 2, Section 2.1). This sample size was adequate for understanding how carriage varied in a given child over the chosen period of time; however, to understand carriage dynamics of the entire population, analysing a larger dataset would be recommended. The overall aim of this study was to understand the transmission of pneumococcal strains in carriage. Although sampling was conducted every four weeks from birth up to one year (Chapter 2, Section 2.1), this study only analysed samples from four subsequent sampling points, previously shown to be positive for *S. pneumoniae* in each subject. The DNA was isolated (Chapter 2, Section 2.5) and used to identify pneumococcal serotype(s) at each sampling point by microarray (Chapter 2, section 2.6). Whole genome sequencing (Chapter 2, section 2.7) was performed on each serotype to determine the genetic diversity. Prior to microarray serotyping, pneumococcal serotypes at each sampling point were initially identified using Quellung, which was employed as a routine serotyping technique commonly used in a clinical laboratory set up. This provided an opportunity to define and compare the serotypes identified at each sampling point by the two serotyping methods.

4.1. Results

4.1.1. Serotype concordance between microarray and Quellung reaction

Sixty-three (63) pneumococcal strains were detected; however only 52 were successfully serotyped using Quellung to yield 24 distinct serotypes (Figure 4.1). The serotypes of the remaining 11 strains could not be characterised, and were thus designated as unidentified (UD). While nontypeable serotypes could not be identified by Quellung reaction, it must be pointed out that the full panel of antisera for Quellung reaction was not available. This may have contributed to 'UD' serotypes, and not necessarily failure by Quellung to identify serotypes. The microarray detected 102 pneumococcal strains, which was 62% higher detection rate compared to Quellung. The microarray provided a serotype on all the pneumococcal strains, yielding 26 distinct serotypes and nontypeables (Figure 4.1). These findings suggest that the microarray was able to detect more pneumococcal strains and was able to ascribe a serotype to each strain detected. The inflated number of serotypes was attributed to the detection of multiple carriage by the microarray. Conventional serotyping by the Quellung reaction was employed as a routine serotyping technique commonly used in a clinical laboratory set up and there was no attempt to determine multiple carriage by this technique at the time of initial sampling. The requirements to (i) undertake multiple Quellung reactions to determine serotype and (ii) have a full set of antisera to complete this process make it a hugely unfeasible process particularly for large scale carriage studies.

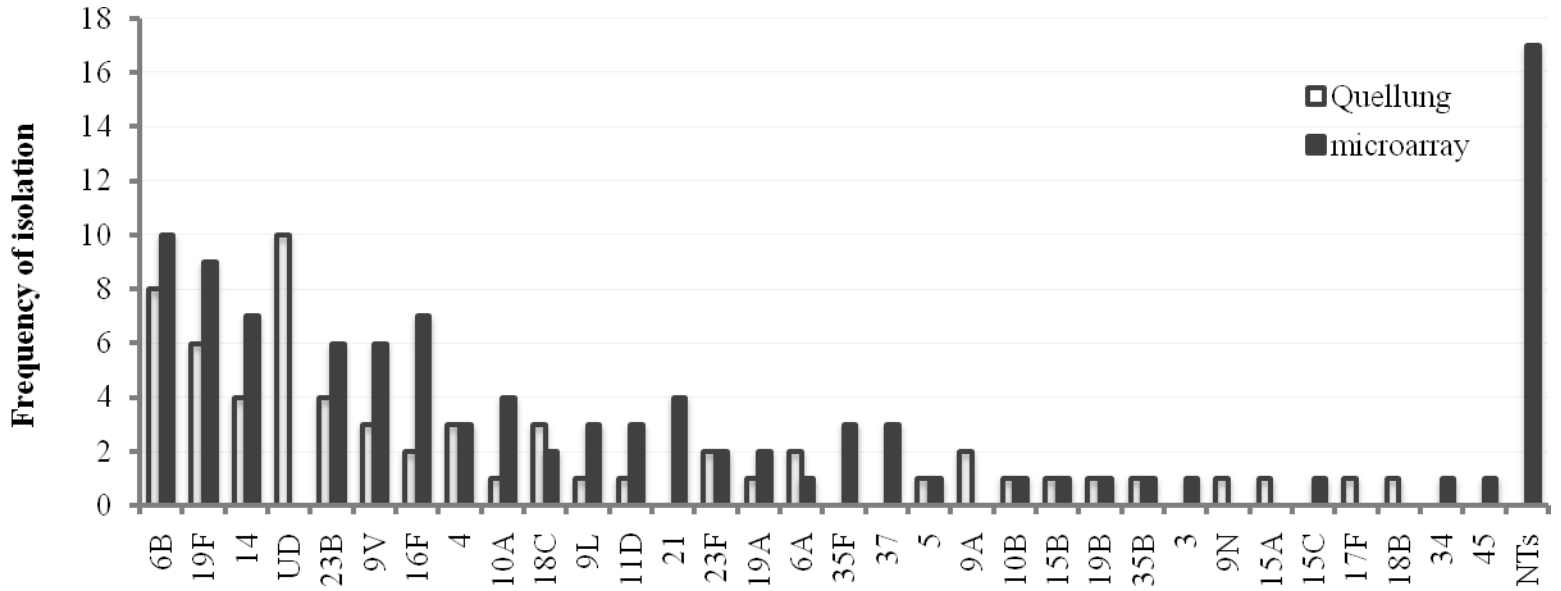


Figure 4.1 Serotypes detected by microarray and Quellung reaction

For this cross-sectional data analysis, only samples from the first sampling point were used. The data showed a wider range of serotypes detected using microarray serotyping compared to the Quellung reaction.

To determine the concordance between the two serotyping methods, only samples for which a serogroup/serotype was assigned by both methods were compared. A common serogroup type was identified in 91% (48/52) of all pneumococcal strains, which is in agreement with previous findings (Turner et al., 2011). The agreement by serotype was 70% (37/52), which was lower than serogroup concordance suggesting the difficulty by the two methods to resolve pneumococcal serotypes particularly within a given serogroup because they are closely related. To determine the accuracy of detection by Quellung only the most abundant serotype in multiple carriage events were considered. Results showed that Quellung reaction detected the most abundant serotype in 67% (22/33) of multiple carriage events, demonstrating that Quellung serotyping was likely to detect the most abundant serotypes in multiple carriage as opposed to the least abundant ones (Table 4.1). This is attributed to the fact that in a routine clinical laboratory set-up, only a single colony is picked for serotyping by the Quellung, thus the most abundant serotype has a higher probability of being sampled for analysis. With such a bias, Quellung reaction most underestimate the number and serotype profiles in circulation, which may have negative implications for assessing the impact of vaccination as well as understanding transmission of pneumococcal strains within the community.

Table 4.1 Serotyping by microarray and Quellung reaction

Sample	Sampling point	Microarray (Relative abundance)	Quellung
A	1	*19F (95%) + NT4a (5%)	19F
	2	*19F (99%) + 9V (1%)	19F
	3	*9V (65%) + 19F (35%)	9A
	4	*9V (94%) + 19F (3%) + 37 (3%)	9V
B	1	*16F (99%) + NT4a (1%)	16F
	2	*9V (95%) +NT4b (3%) + 37 (2%)	9V
	3	NT4b (55%) + 37 (45%)	18B
C	1	*14 (54%) + 23B (46%)	14
	2	*23B (94%) + 14 (6%)	23B
D	1	NT3b (41%) + 35B (37%) + NT2 (22%)	35
	2	*19F (99%) + NT4a (1%)	19F
E	1	*19F (93%) + 35F (7%)	19F
	2	35F (58%) + 19F (42%)	UD
	3	35F (95%) + 19F (5%)	15A
F	1	*15B (98%) + 16F (2%)	15B
	2	*15C (87%) + 3 (13%)	15B
	3	6B (76%) + 16F (24%)	UD
	4	*6B (82%) + NT3b (18%)	6B
G	1	*11D (85%) + NT4a (15%)	11
	2	*19A (89%) + 11D (11%)	19A
H	1	*16F (60%) + NT3b (20%) + NT2 (20%)	16F
	2	*9V (59%) +16F (41%)	9V
I	1	*10B (51%) + NT4b (44%) + 34 (5%)	10B
J	1	9L (91%) + 23B (9%)	23B
	2	*23B (95%) + 10A (5%)	23F
	3	*9L (85%) + 23B (5%)	9N
K	1	16F (51%) + 14 (49%)	17F
	2	NT3b (67%) + NT3a (33%)	UD
	3	NT3B (89%) + 14 (11%)	UD
L	1	21 (68%) + 6B (33%)	6B
	2	*4 (80%) + 21 (20%)	4
	3	*23F (99%) + 4 (1%)	23B
	4	*23F (74%) + NT3b (26%)	23F

Strains detected by Quellung reaction were present as the most abundant (*) in multiple carriage events by microarray. (UD) denotes unidentified serotypes.

4.1.2. Pneumococcal carriage dynamics over time by microarray

The carriage dynamics of specific pneumococcal serotypes over time was investigated using longitudinal samples from a subset of children (n=16) in Karonga. The aim was to establish whether the ability by microarray to detect low abundance pneumococcal strains would extend the measured carriage duration of pneumococcal strains and whether there was any difference by serotype. A total of 63/64 (one sample did not grow on subsequent culturing) samples collected at four sequential time points from each of the 16 children were analysed. Serotyping by microarray yielded 26 distinct serotypes, including nontypeablestrains and at each sampling point, at least one strain was detected. Quellung reaction only identified one serotype at each sampling point and this was largely the most abundant serotype in multiple carriage events.

Carriage dynamics of single serotypes over time. Microarray results showed that 25% (4/16) of the children carried a single serotype at each sampling point (Table 4.2 A). These children were C3, C9, C12, and C13.Three of the 4 children (C3, C9, C13) carried the same serotype at three subsequent sampling points (Table 4.2 A and B) and a different serotype was detected at each sampling point for the remaining child (C12) (Table 4.2 C). These data indicate that over time, carriage of a particular serotype persisted in a given individual, while in other cases; different serotypes were carried at each sampling point highlighting the complexity of carriage.

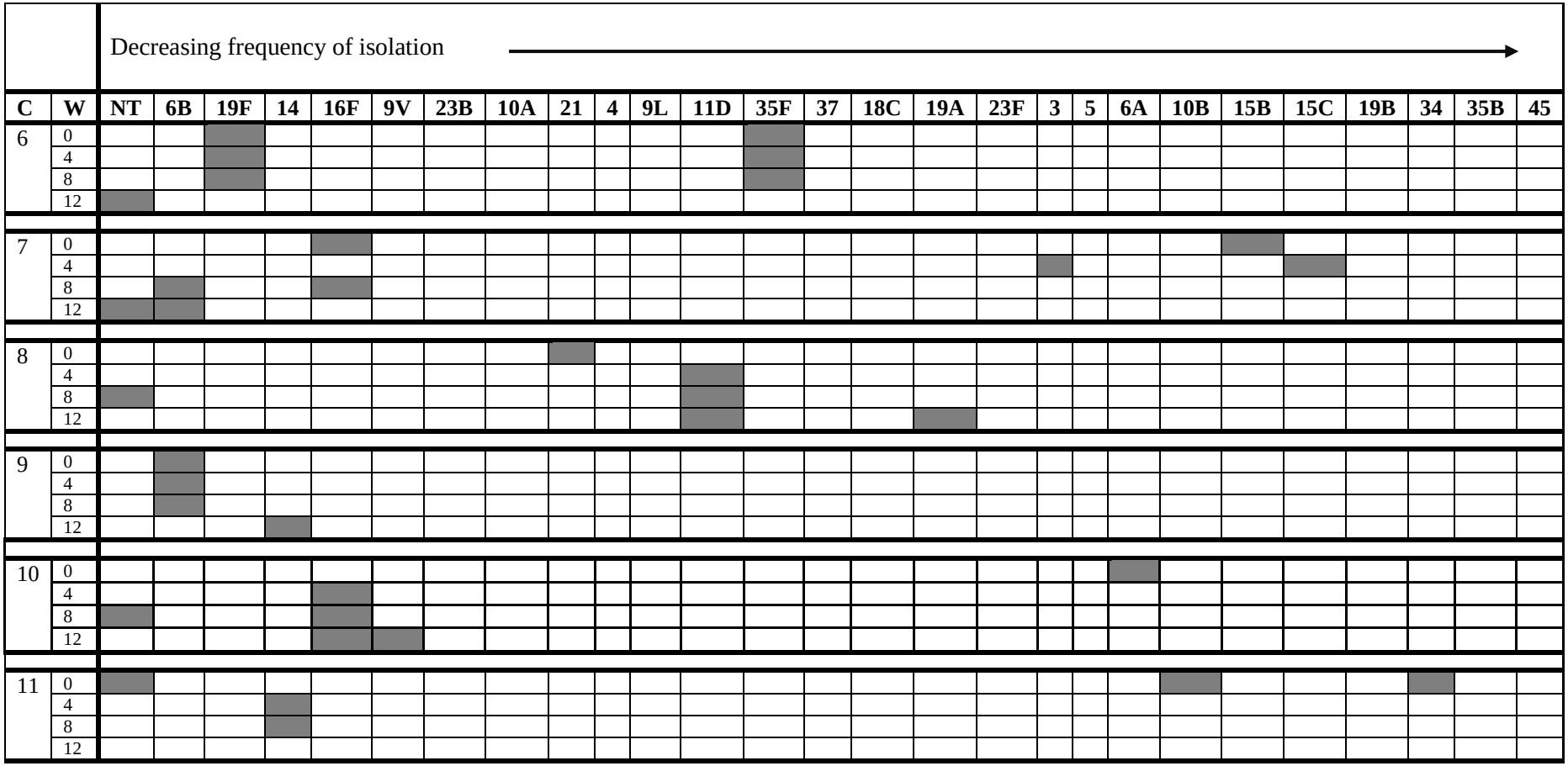
Table 4.2 Serotype specific carriage dynamics overtime by microarray

A

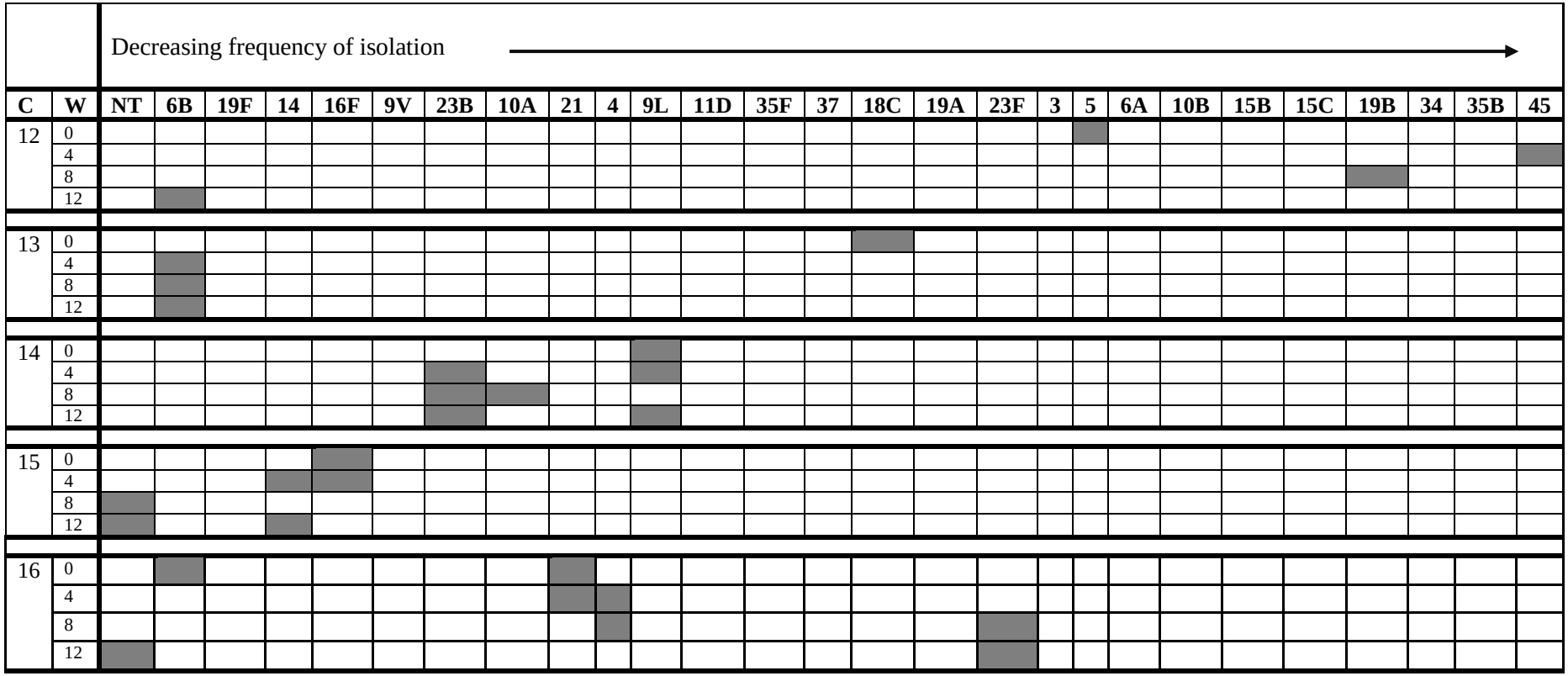
		Decreasing frequency of isolation →																										
C	**W**	**NT**	**6B**	**19F**	**14**	**16F**	**9V**	**23B**	**10A**	**21**	**4**	**9L**	**11D**	**35F**	**37**	**18C**	**19A**	**23F**	**3**	**5**	**6A**	**10B**	**15B**	**15C**	**19B**	**34**	**35B**	**45**
1	0	■		■																								
	4			■			■																					
	8			■			■																					
	12			■			■								■													
2	0	■				■																						
	4						■								■													
	8	■					■								■													
	12	■													■													
3	0								■																			
	4								■																			
	8								■																			
	12																■											
4	0															■												
	4				■			■																				
	8							■																				
	12				■			■																				
5	0	■																									■	
	4			■																								
	8	■		■																								
	12										■																	

The samples were collected from 16 children (C) every four weeks (W). The shaded regions denote carriage of a given pneumococcal serotype at each sampling point. The data is divided into three tables: A, B, and C

B


Decreasing frequency of isolation
C
W
NT
6B
19F
14
16F
9V
23B
10A
21
4
9L
11D
35F
37
18C
19A
23F
3
5
6A
10B
15B
15C
19B
34
35B
45
6
0
4
8
12
7
0
4
8
12
8
0
4
8
12
9
0
4
8
12
10
0
4
8
12
11
0
4
8
12

C


Decreasing frequency of isolation
C
W
NT
6B
19F
14
16F
9V
23B
10A
21
4
9L
11D
35F
37
18C
19A
23F
3
5
6A
10B
15B
15C
19B
34
35B
45
12
0
4
8
12
13
0
4
8
12
14
0
4
8
12
15
0
4
8
12
16
0
4
8
12

Carriage dynamics of multiple serotypes over time. Out of the 16 children, 12 were shown to carry multiple serotypes at one sampling point (n=1), two sampling points (n=4), three sampling points (n=4) and all the four sampling points (n=3) (Table 4.2). This represented 75% (12/16) multiple carriage, however based on the total number of samples (n=63) from each sampling point, multiple carriage was detected in 33 samples which represented 52 % (33/63) multiple carriage rate (Section 4.1.4). The relative abundance (%) of each of the serotypes in multiple carriage was measured based on fluorescence intensity on microarray. This allowed the progression of serotypes in carriage over time to be tracked. For example, serotypes 19F and 35F (Figure 4.2) were detected at the first three sampling points (wk0 wk4, and wk8).

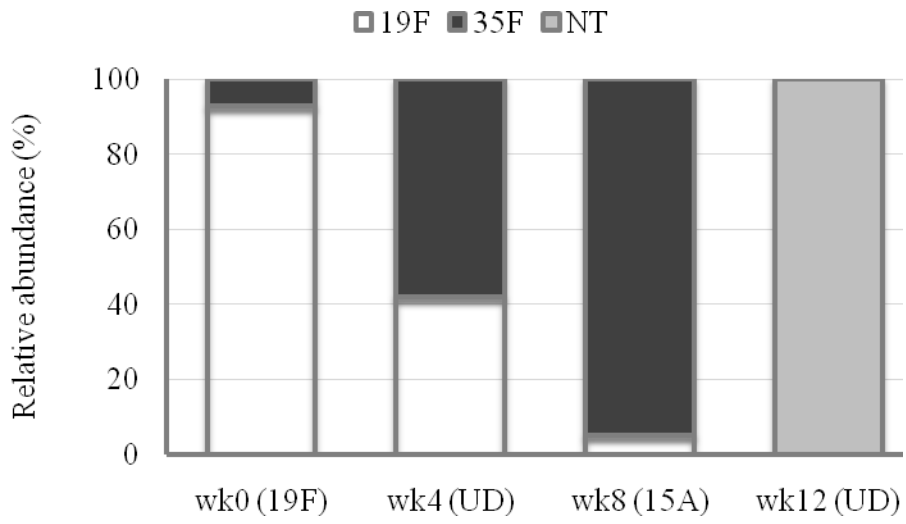


Figure 4.2 Dynamics of multiple serotype carriage over time

Showing carriage of two serotypes at each sampling point with changing relative abundance (%) over time (weeks). In brackets are serotypes detected by Quellung method. Pneumococcal strains that could not be serotyped by Quellung were designated UD (undetermined).

The relative abundance for serotype 19F decreased from 93% to 5%, while 35F increased from 7% to 95% within the same time period. None of the two serotypes (19F and 35F) were detected at the fourth sampling point except for a nontypeable serotype (NT). Figure 4.3 demonstrates carriage of serotype 19F across all four sampling points, with decreasing relative abundance over time.

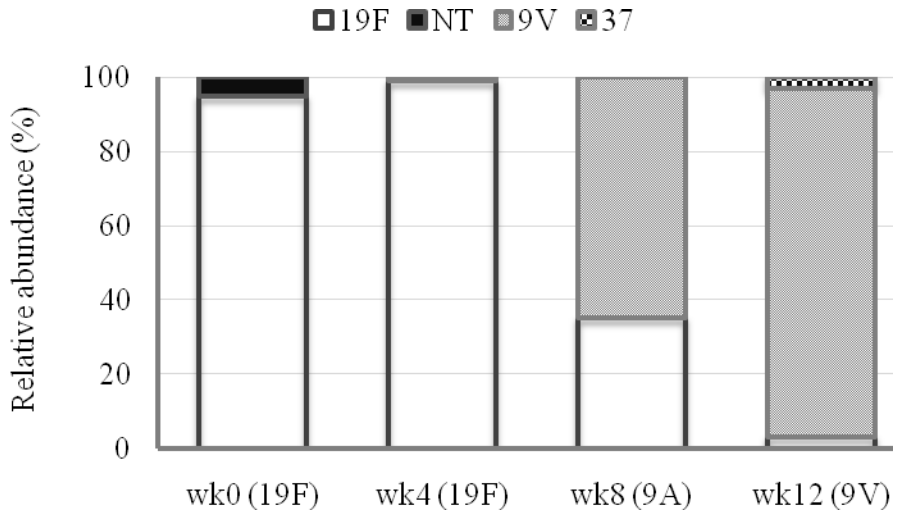


Figure 4.3 Dynamics of multiple serotype carriage over time

Showing carriage of two or three serotypes at each sampling point with changing relative abundance (%) over time (weeks). In brackets are serotypes detected by Quellung method.

At the first sampling point (wk0), serotype 19F existed as the most abundant serotype (95%) in multiple carriage with nontypeable serotype (NT) representing the remaining 5%. Across the three further sampling points, 19F remained but with a concurrently increasing relative abundance of serotype 9V. The relative abundance of serotype 19F decreased from 99% to 3%, while serotype 9V went from a low 1%

relative abundance to become the most abundant serotype at 94%. An additional serotype, serotype 37, at 3% relative abundance was detected at the last sampling point. The results above show a time dependent variation in relative abundance, which could be attributed to competition among colonising bacteria or evolving host immune response. Using the microarray, the acquisition of new serotypes was well demonstrated, for example, serotypes NT and 37 (Figures 4.2 and 4.3) were only present at the last sampling point although they were not observed at the previous three sampling points.

4.1.3. Genetic diversity of pneumococcal serotypes over time

4.1.3.1. Whole genome sequencing

Analysis by whole genome sequencing was performed only on pneumococcal isolates detected at more than one sampling point. The aim was to determine whether or not, there were any genetic similarities between isolates from subsequent sampling points, given the high rates of genetic transformation that the pneumococcus undergoes, which sometimes occurs within a very short period of time (Calix et al., 2012a), genetic analysis would determine if this had occurred. The pneumococcal serotypes were isolated from a re-culture of the original sample. Single colonies were picked and inoculated into BHI broth. The bacteria were harvested and DNA extracted as indicated previously (Chapter 2, section 2.5). To confirm the serotype at each sampling point, the DNA from each sample was subjected to PCR analysis by amplifying serotype specific genes within the capsular polysaccharide locus (Pai et al., 2006) (Appendix 7a). The PCR amplicons were analysed on a Bioanalyser (Chapter 2, section 2.5), which generated fluorescence units (FU) vs base pair (bp) length graphs (Electropherograms) of the respective amplified genes for each

serotype (Figures 4.4) (Appendix 7b). Serotypes were positively identified by the presence of a band with the expected size on a gel. For example, the identity of serotype 9L was established through the amplification of the *wzx* gene with an expected size of 516 base pairs. Serotypes 11D and 19F were both identified by amplifying wzy genes but with expected sizes of 304 and 463 base pairs respectively (Figure 4.4).

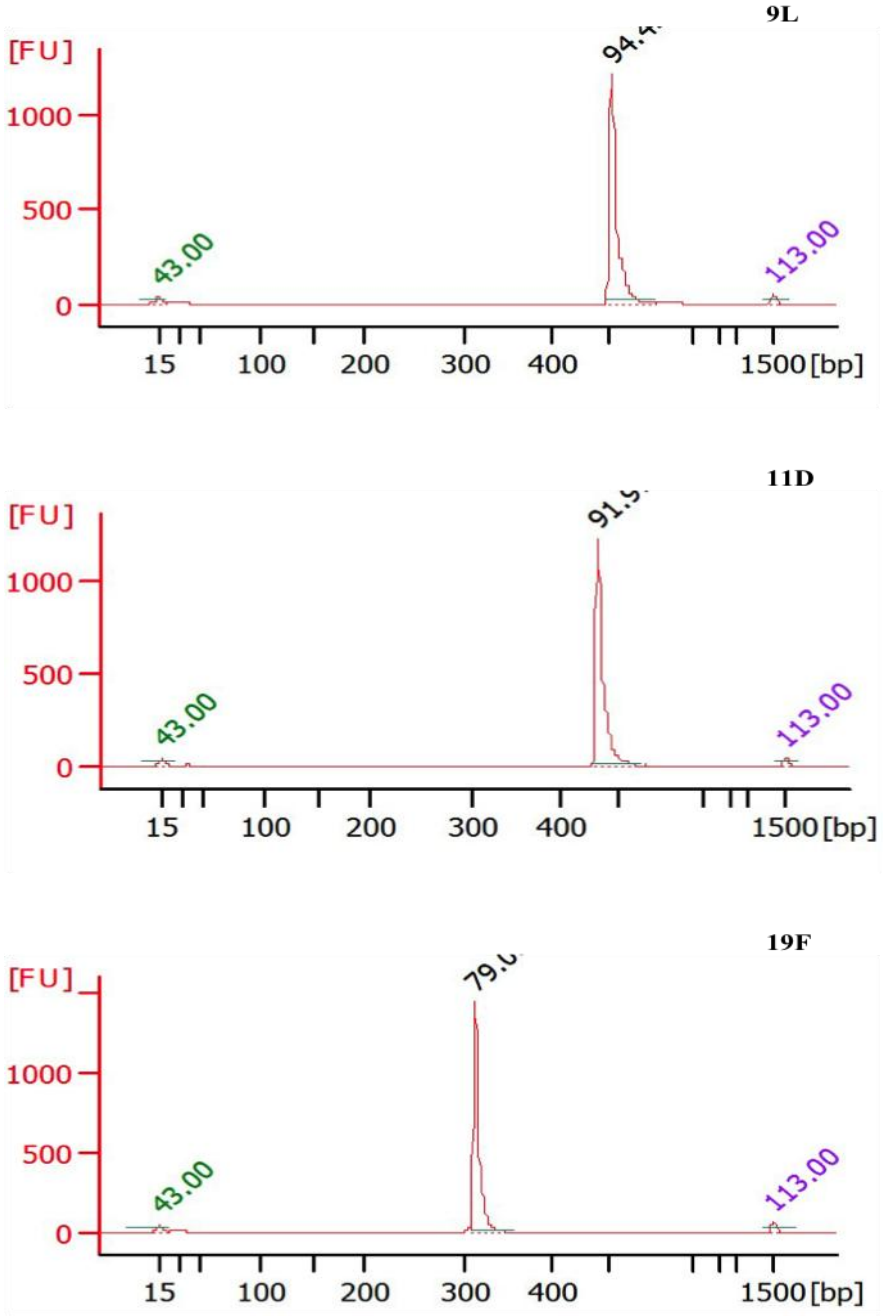


Figure 4.4 Bioanalyser electropherograms

This confirmed the presence of 9L (512bp, *wzx*), 11D (466bp, *wzy*) and 19F (311bp, *wzy*), as previously determined by microarray. Similar electropherograms (Appendix 4b) were generated for 6B (250bp, *wciP*), 9V (816bp, *wzy*), 16F (716bp, *wzy*), 23B (199bp, *wzx*), 23F (384bp, *wzy*), and 35F (517bp, *wzy*) (Pai et al., 2006, da Gloria Carvalho et al., 2010, Dias et al., 2007).

4.1.3.2. Multi-locus sequence typing

MLST was performed on whole genome sequences using an in-house script developed at the Sanger Institute in Cambridge (UK). The aim was to establish whether subsequent serotypes belonged to the same or different genotypes, which in-turn would inform the dynamics of carriage in the children over time. The genotype of each serotype was defined based on the MLST allelic profile obtained from the MLST database (http://www.spneumoniae.mlst.net) (Table 4.3). Thirty-seven pneumococcal strains were analysed: 6B (x5), 9L (x3), 9V (x2), 10A (x3), 11D (x3), 14 (x2), 16F (x3), 19F (x8), 23B (x3), 23F (x2), and 35F (x3). The MLST allelic profiles for 90% (28/37) of the pneumococcal strains analysed were successfully obtained. Of these, 50% (14/28) allelic profiles not reported in the database, and their sequence types were thus designated as 'novel'. The sequence types for the remaining 14 samples were retrieved from the database: 10A (x3) (ST 5483), 14 (x2) (ST 63), 19F (ST 6095 (x5) and ST 347 (x2) and 23F (x2) (ST 802). In 9 pneumococcal samples, the sequence types were unknown because the allele number for at-least one locus could not be established; hence they were reported as 'nd' (not determined) (Table 4.3). The allelic profile for 6B serotypes isolated from two different children matched at only one (*xpt*) of the seven loci, while in serotype 19F, 2 of the three children carried sequence type 6095 at all sampling points with the remaining child carrying sequence type 347. This suggested that more than one genotype of 6B and 19F were in circulation. The data showed that a particular individual carried the same pneumococcal genotype over the entire sampling period, however genotype variations were observed among different carriers of the same serotype. There is a possibility the genotype could change if carriage persisted for a much longer period of time, however this would require further investigation.

Table 4.3 Multilocus sequence typing

A

Sample Accession #	Sero-type	Sequence-type	Cluster	*MLST allelic profile*								
				AroE	***gdh***	***gki***	***recP***	***spi***	***xpt***	***ddl***	***Closest ST match***	***Notes***
ERS67710	6B			1	43	29	1	77	1	14		Double locus variant at *gdh* and *spi* alleles,
ERS67708	6B	Novel-1	C3	1	43	29	1	77	1	14	8783	isolated from carriage NT in china
ERS67709	6B			1	43	29	1	77	1	14		
ERS67687	6B	Novel-3	C8	2	8	1	10	17	1	19	2863; 4441;	Double locus variant (*gki/ddl*) alleles (2863)
ERS67688	6B			2	8	1	10	17	1	19	5266	and gdh/gki (4441/5266) alleles
ERS67700	9L			2	12	8	8	27	28	5	4914	Double locus variant at gdh/spi alleles (4914, 9L)
ERS67701	9L	Novel-5	C10	2	12	8	8	27	28	5	8833	Andgdh/xpt alleles (8833, 9L)
ERS67702	9L			2	12	8	8	27	28	5		
ERS67681	9V	nd	C14	nd	5	36	3	6	1	6		
ERS67680	9V			nd	5	36	3	6	1	6		
ERS67677	10A			8	11	255	47	6	346	9		Four different strains isolated from blood in Mozambique
ERS67679	10A	5483	C1	8	11	255	47	6	346	9		
ERS67678	10A			8	11	255	47	6	346	9		
ERS67699	11D			2	18	6	10	17	20	19	5266	Single locus variant at *xpt* allele (22A), isolated from
ERS67698	11D	Novel-2	C7	2	18	6	10	17	20	19		bacteremia in Kenya
ERS67697	11D			2	18	6	10	17	20	19		
ERS67703	14	63	C6	2	5	36	12	17	21	14		Similar ST present in serotypes: 15A, 23F,
ERS67704	14			2	5	36	12	17	21	14		19A, 19F and 23A from Europe and USA

B

Sample Accession #	Sero-type	Sequence-type	Cluster	*MLST allelic profile*								
				AroE	***gdh***	***gki***	***recP***	***spi***	***xpt***	***ddl***	***Closest ST match***	***Notes***
ERS67705	16F			50	5	1	nd	6	88	14		
ERS67706	16F	nd	C4	50	5	1	nd	6	88	14		
ERS67707	16F			nd	nd	nd	nd	nd	nd	nd		
ERS67675	19F			12	8	9	104	3	20	417		Isolated from bacteraemia in Kenya
ERS67674	19F	6095	C11	12	8	9	104	3	20	417		
ERS67676	19F			12	8	9	104	3	20	417		
ERS67689	19F	6095		12	8	9	104	3	20	417		Isolated from bacteraemia in Kenya
ERS67692	19F	nd	C12	nd	nd	nd	nd	nd	nd	nd		
ERS67694	19F	6095		12	8	9	104	3	20	417		
ERS67682	19F	347	C13	12	8	9	3	3	20	57		Isolated from blood in South Africa and carriage in Norway
ERS67683	19F			12	8	9	3	3	20	57		
ERS67686	23B			2	43	54	5	7	20	6	5419	Double locus variant (gdh/recP, 6B) from blood in Mozambique
ERS67684	23B	Novel-4	C9	2	43	54	5	7	20	6	6078	Triple locus variant (gdh/spi/xpt, 6A) from bacteraemia in Kenya
ERS67685	23B			2	43	54	5	7	20	6	7530	Triple locus variant (gdh/recP/spi, 34) from carriage in Nepal
ERS67712	23F	802	C5	10	13	53	1	72	38	31		Similar ST present in serotypes: 19F (China), 5 and 19A (Gambia)
ERS67711	23F			10	13	53	1	72	38	31		
ERS67690	35F			8	11	255	47	13	346	nd		
ERS67691	35F	nd	C2	nd	nd	nd	nd	nd	nd	nd		
ERS67693	35F			8	11	255	47	13	346	nd		

Sub-tables A and B show genotyping by MLST.

4.1.3.3. Phylogenetic analysis

A phylogeny is used to describe the evolutionary relationship among various biological species based on their physical or genetic properties. In this study, a phylogenetic analysis was performed on all subsequent serotypes to establish whether serotypes isolated from different sampling points were genetically related. This would consequently provide information on whether carriage at each sampling point over time involved the same strain or a different strain. The tree was generated based only on SNPs of whole genomes (Chapter 2, Section 2.9.1), using the genome of serotype 23F (FM211187) as standard reference (Figure 4.5). The sample accession numbers were used to identify respective serotypes at each sampling point. The serotypes appear in brackets immediately after the accession numbers. The tree generated 14 clusters (C1 to C14) (Figure 4.5). Each cluster was linked to the serotype and the child providing the samples. Serotypes contained in clusters C2 (35F) and C12 (19F) were isolated from a single child. Serotype 6B showed two clusters, each containing two (C8) and three (C3) samples respectively. It was determined that clustering in 6B serotypes was linked to two different children providing the samples. Similarly, 19F serotypes belonged to three different clusters corresponding to three different children who provided the samples. This showed that similar serotypes isolated from different subjects were not phylogenetically related, and this was prominent in serotype 6B. However, similar serotypes, isolated from the same child, were phylogenetically related and belonged to the same cluster. Further, each cluster was shown to contain serotypes, which belonged to the same genotype (Table 4.2). These findings suggest two things: (i) there is high genetic diversity of carriage serotypes in Malawian children and (ii) the same serotypes isolated from subsequent sampling points in a given child over time belonged to the

same phylogenetic cluster suggesting they were genetically similar; a finding supported by MLST genotyping (Table 4.2).

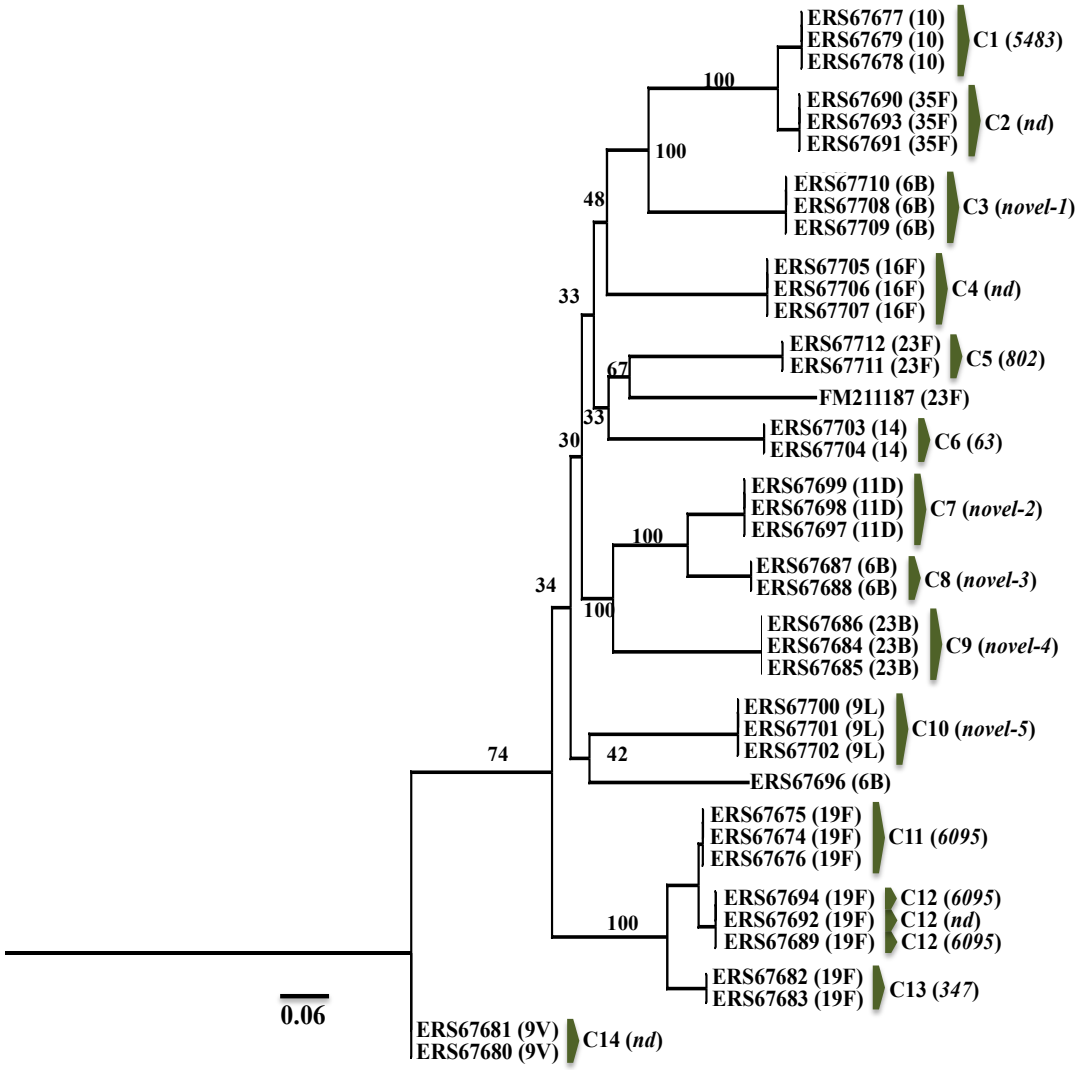


Figure 4.5 Maximum likelihood phylogenetic tree

The tree was generated using the SNPs of the whole genome of *S. pneumoniae.* The green arrows on the right indicate strains belonging to the same sequence type (in brackets), which were either known, novel, or not determined as defined before (Section 4.6.2).

4.1.3.4. Genetic distance between subsequent serotypes

Genetic distance is a measure genetic difference between closely related species (Nei, 1972). To determine genetic distance, pairwise genetic distance (d) between strains was calculated as a Hamming distance. Hamming distance compares each nucleotide base in aligned genomic sequences of two sequences selected from the generated multiple sequence alignment. For example, the genetic distance between ACGTTACGA and ATGCTATGA is 3. In the present study, the genetic distance was determined (Figure 4.6), to establish whether subsequent serotypes at each sampling point belonged to the same or different strains.

For clusters containing two serotypes (Figure 4.5), the comparison yielded a single value for SNPs differences as there could only be one comparison made: C4_16F (5 SNPs); C5_23F (5 SNPs); C6_14 (3 SNPs); C8_6B (12 SNPs); C13_19F (16 SNPs) and C14_9V (11 SNPs). In brackets are values for the SNPs differences between serotypes at two subsequent sampling points. However, for all the clusters containing three serotypes (C1, C2, C3, C7, C9, C10, C11, C12), three possible SNPs comparisons were made, yielding a maximum number of three SNPs values. The results (Figure 4.6) showed SNPs differences existed between subsequent serotypes; however these differences were not statistically significant, suggesting that at each subsequent sampling point, each child carried the same genotype from the same strain.

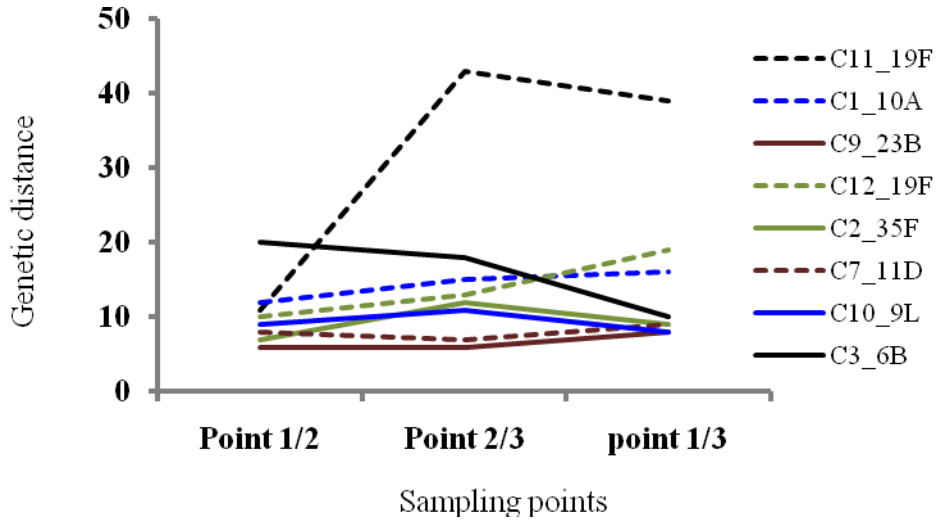


Figure 4.6 Genetic distance based on SNPs

Genetic distance was derived for serotypes from subsequent sampling points. Each graph represents a particular serotype from a given phylogenetic cluster (Figure 4.5): C1, C2, C3, C7, C9, C10, C11, C12.

A comparison of SNPs in serotypes with the same genotype but isolated from different carriers, showed a much higher difference in genetic distances compared to those from the same carrier regardless of similarities in genotype (Table 4.5). This would suggest carriage of different strains in each infant belonging to the same geographical area.

Table 4.3 Genetic distance based on whole genome SNPs

		Genetic distance	
Serotype	**Cluster**	**Sampling point 1**	**Sampling point 2**
	C11/C13	5395	5386
19F	C11*/C12*	1631	1616
	C12/C13	6346	6408
6B	C3/C8	18912	18277

*Genetic distance was generated for similar serotypes from different carriers. Each cluster shown was from a different infant carrier. *Belonged to the same genotype.*

4.2. Discussion

This chapter has investigated the sensitivities of using microarray serotyping and the conventional Quellung reaction at detecting pneumococcal serotypes in carriage, using longitudinal samples from children. The data generated relate to the diversity and dynamics of pneumococcal carriage in Malawian children.

Sensitivity of molecular serotyping by microarray compared with Quellung reaction. Pneumococcal serotype distribution was analysed in a longitudinal children's dataset (n=16), using Quellung and microarray serotyping. The microarray identified 62% more serotypes than the Quellung reaction because of multiple carriage detection although the number of distinct serotypes detected was not significantly different (Figure 4.1). Nontypeable serotypes were only detected by microarray, whereas the Quellung reaction reported several unidentified serotypes. Serogroup concordance between the Quellung reaction and microarray platforms, across each sampling point was 91%, a high level of agreement, similar to a previous report on concordance between latex agglutination and microarray serotyping, where a 95.2% concordance by serogroup was reported (Turner et al., 2011). However the endpoint in most carriage studies is to estimate vaccine coverage, at the level of serotype. An investigation into serotype concordance showed a 70% (37/52) agreement between the two methods with microarray detecting several additional serotypes not identified by Quellung reaction. In addition, Quellung reaction was biased towards detecting the most abundant serotype in samples containing multiple serotypes, leaving the less abundant serotypes unreported. These data suggest Quellung reaction underestimates the profile of serotypes in circulation and inadequate for providing a good estimate of vaccine coverage in a given population.

Although Quellung reaction has a number of limitations, as previously discussed, it is still regarded as the gold standard (O'Brien and Nohynek, 2003) and has provided good data on serotypes prevalent in invasive disease in various settings (Johnson et al., 2010) including Malawi (Cornick et al., 2011). However, it still lacks the required sensitivity to detect serotypes present in low abundance in carriage, which is key to understanding the differences in serotype distribution in carriage and disease and to effectively monitor the impact of vaccine implementation.

Dynamics of pneumococcal carriage over-time. This study has shown that carriage of pneumococcal serotypes is very diverse in the Malawian setting with a high prevalence of multiple carriage (Chapter 3, Section 3.3 and 3.5). Overall results suggested no change in serotype distribution over the entire sampling period. However, the frequency with which pneumococcal strains belonging to different serotypes were isolated from subsequent sampling points was high for carriers of both single and/or multiple serotypes; suggesting high level of pneumococcal transmission. The complexity of carriage dynamics was further demonstrated in a large number of multiple serotype carriers, where the abundance of one serotype was seen to increase with a simultaneous decrease in the abundance of other co-colonising serotypes (Figures 4.2 and 4.3). This could be attributed to competition among co-colonising serotypes, which shows an important role multiple carriage plays in regulating the distribution of pneumococcal serotypes in carriage. The maturation of the host immunity overtime is also another factor that regulates carriage dynamics (Bogaert et al., 2009, Cobey and Lipsitch, 2012).

Genetic diversity of pneumococcal serotypes carried over-time. The genetic diversity over time was investigated by analysing whole genome sequences of all serotypes that were isolated from sequential sampling points. Altogether sequences for 37 pneumococcal strains (11 distinct serotypes) from 13 children were analysed. MLST analysis showed that pneumococcal strains belonging to different serotypes had different genotypes, and there was no evidence of capsule switch observed. Capsule switch is reported to have occurred in a particular setting if two different serotypes express the same genotype. However the genotype of pneumococcal strains belonging to the same serotype but isolated from different children was inconclusive (Table 4.2). For example serotype 19F was isolated from three different children, two children carried ST6095 while the remaining child carried ST347. The only two 6B serotypes sequenced in this section showed they belonged to two different genotypes, highlighting the previously reported genetic diversity in serogroup 6 (Elberse et al., 2011). Two different genotypes were also detected in 6B serotypes obtained from two different infants. These data demonstrate the high diversity of pneumococcal genotypes circulating in Malawi. The sequence type of pneumococcal strains belonging to the same serotype and isolated from the same infant at sequential sampling points remained unchanged. To confirm whether the infants carried the same strain at each sampling point, the genetic distance based on SNPs differences was compared (Section 4.1.3.4). The analysis showed no major differences in the genetic distance for each serotype at each sampling point suggesting that over-time; the child carried the same strain (Figure 4.6). However, the difference in the genetic distance between similar serotypes (including those with same genotype) from different carriers was high (Table 4.5), which suggested carriage of different strains.

4.3. Conclusion

We have demonstrated that the serogroup concordance between microarray and was high (93%) than their agreement on serotype level (70%). We have also demonstrated, by microarray, that pneumococcal carriage is as diverse as it is dynamic in Malawian children. Thus depending on the objectives of a particular study, microarray though relatively costly, would provide more accurate information on carriage dynamics, consequently benefiting the assessment of vaccine impact. We have also shown that serotypes in circulation are genetically diverse; however despite high levels of multiple carriage, there was no evidence of capsule switch in this cohort.

Chapter 5: Characterisation of capsular polysaccharide locus variants

5.0. Introduction

The capsular polysaccharide biosynthetic genes are located within the CPS locus, flanked by*dexB* and *aliA* genes. The polysaccharide capsule defines serotype, and based on the pattern of reactivity with the factor typing sera, 95 individual serotypes are currently recognised (Park et al., 2007, Bratcher et al., 2010, Oliver et al., 2013). The capsule is a target for the host antibodies and has been shown to maintain the survival of the pneumococcus in the host by acting as a barrier to the host killing by phagocytosis. The capsule is central to the design of pneumococcal conjugate vaccines. Under vaccine selection pressure, or indeed naturally, the pneumococcus may undergo genetic changes within the CPS locus to maintain its survival within the host. Thus genetic changes within the CPS locus may alter the capsular polysaccharide biochemical structure, which could lead to capsule change or switch consequently impact on pneumococcal virulence and response to treatment strategies. In this study, molecular serotyping was performed by microarray which has the enhanced ability to detect novel serotypes (Newton et al., 2011). Microarray detected sequence divergence and absent or highly divergent genes within the CPS locus of some serotypes. These were designated CPS locus variants. Whole genome sequencing was performed on such serotypes to identify the genetic changes that might have occurred at the CPS locus. This chapter presents a genomic analysis of the subset of serotype variants detected by microarray in the entire data set. Specifically, the analysis focused on genetic variations at the CPS locus of whole genomes of types 6B and 19A, which are covered by the current conjugate vaccine

(PCV13) and have been isolated from invasive disease in Malawi (Cornick et al., 2011) and non-vaccine types 9l, 15B, 20, 23B and 29.

5.0.1 Methods

The detailed methods for this section are outlined in Chapter 2, sections 2.7 to 2.9.5

5.0.2 Results

5.1. Analysis of the CPS locus variants by phylogeny and MLST

The phylogenetic relatedness of the *CPS* locus variants, together with previously characterised Malawian invasive serotypes was determined. The comparison was based on SNPS across the whole genome, mapping the sequences to the reference serotype 23F strain ATCC700669 (FM211187). The tree was generated using RAxML (Stamatakis et al., 2005) by performing 100 bootstraps. Results showed that phylogenetic clustering was not dependent on sero-group (Figure 5.1) alone, however, all serotypes belonging to a particular sero-group and sharing MLST alleles at six or seven loci formed a single cluster, which resulted in 6 main clusters (cluster A-F). Therefore clustering was associated with similarities across both sero-group and genotype. The CPS locus variants of serotype 6B and 19A were phylogenetically very diverse as they were interspersed throughout the tree (Figure 5.1).

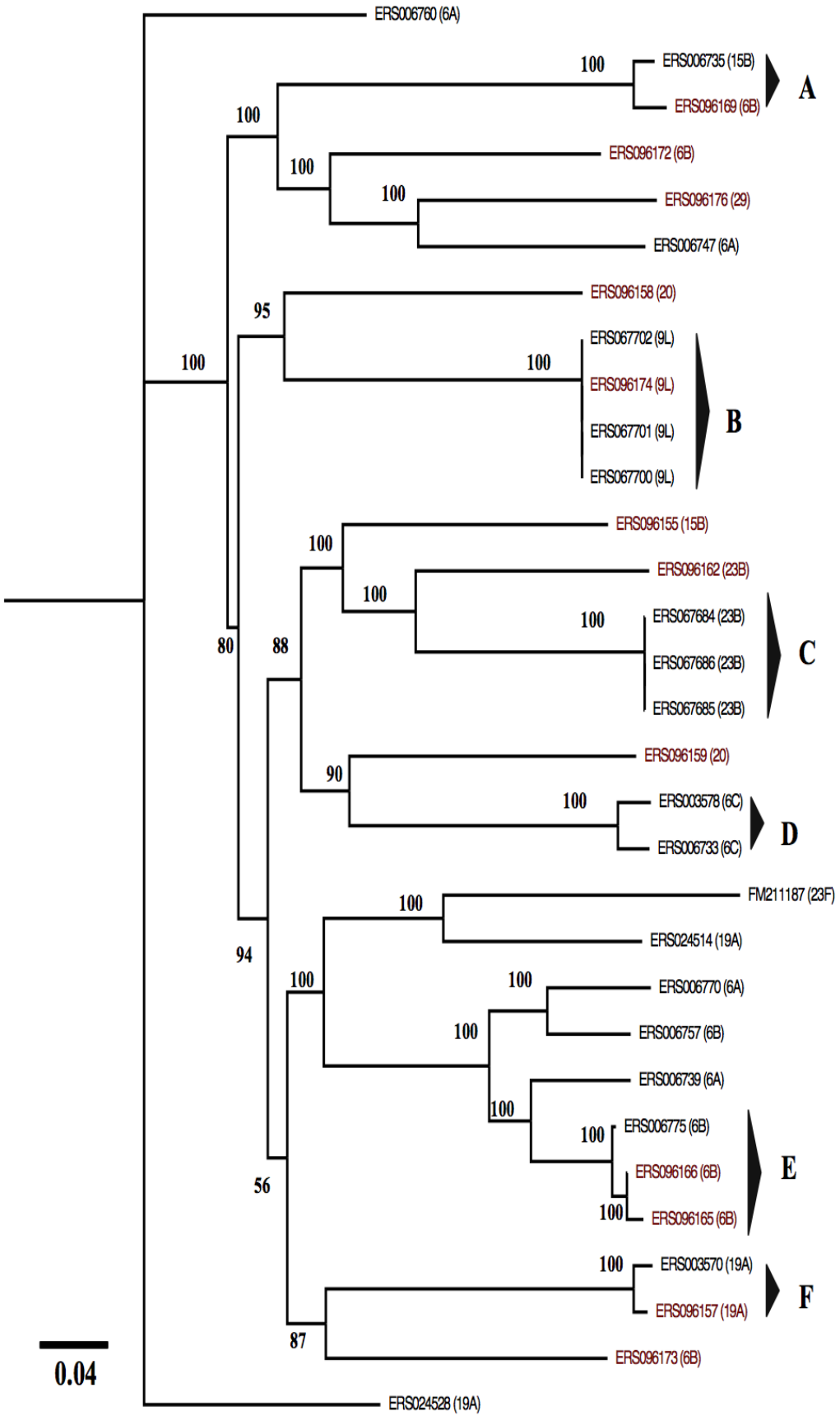

ERS006760 (6A)
ERS006735 (15B)
ERS096169 (6B)
A
ERS096172 (6B)
ERS096176 (29)
ERS006747 (6A)
ERS096158 (20)
ERS067702 (9L)
ERS096174 (9L)
ERS067701 (9L)
ERS067700 (9L)
B
ERS096155 (15B)
ERS096162 (23B)
ERS067684 (23B)
ERS067686 (23B)
ERS067685 (23B)
C
ERS096159 (20)
ERS003578 (6C)
ERS006733 (6C)
D
FM211187 (23F)
ERS024514 (19A)
ERS006770 (6A)
ERS006757 (6B)
ERS006739 (6A)
ERS006775 (6B)
ERS096166 (6B)
ERS096165 (6B)
E
ERS003570 (19A)
ERS096157 (19A)
F
ERS096173 (6B)
ERS024528 (19A)
0.04

Figure 5.1 Phylogenetic analysis CPSlocus variants

Serotypes are shown by their respective sample accession number and the variants are colour coded red. Malawian wild type invasive serotypes (colour coded black) were also included in the tree for comparison. The tree was generated by whole genome alignment of SNPs from each genomic sequence with RAxML, by performing 100 bootstraps. The bootstrap values for each branch are indicated. The black arrows show clusters (A-E) for respective serotypes. The SNPs were identified using reference *Streptococcus pneumoniae* strain ATCC700669 (FM211187).

The two serotype 6B variants (ERS096166 and ERS096165) and a previously characterised invasive 6B (ERS006775) serotype belonged to cluster E. MLST analysis showed that ERS006775 and ERS096166 belonged to the same sequence type while ERS096165 was only a single locus variant of the two serotypes, with a variation at allele number 6 (*xpt* gene) (Table 5.1). A search on MLST database (MLST[database]) revealed that the closest sequence type match for ERS096166 and ERS096165 were ST 6382 and ST 7744 isolated from South Africa, respectively, where alleles at six of the seven loci were shared. Each of the remaining three 6B variants (ERS096173, ERS096172 and ERS096169) exhibited very unique allelic profiles with ERS096173 and ERS096172 matching at only a single *xpt* gene, while ERS096169 did not have any matching allele (Table 5.1). As expected, these three variants were located on different branches of the tree (Figure 5.1). It was further observed that the 6B (ERS096169) variant and wildtype 15B (ERS006735) were phylogenetically related, as they were located on same branch (cluster A). However, MLST analysis showed a single locus variation (*gdh*) between the two serotypes, which could suggest that capsule switch did not occur.

Table 5.1 MLST analysis of CPS locus variants

Strain ID	Serotype	ST	*aroE*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*
ERS006770	6A	2790	8	9	4	13	15	20	5
ERS006739	6A	2285	7	9	4	13	15	20	28
ERS006760	6A	2987	7	13	4	16	6	1	6
ERS006747	6A	Novel	12	5	4	5	6	19	6
ERS006775	6B	Novel	10	9	4	5	15	20	28
ERS006757	6B	Novel	97	9	4	13	15	20	14
ERS096173	6B	Novel	2	8	1	10	17	1	19
ERS096172	6B	Novel	1	43	29	1	77	1	14
ERS096166	6B	Novel	10	9	4	5	15	20	28
ERS096165	6B	Novel	10	9	4	5	15	4	28
ERS096169	6B	Novel	54	5	4	5	36	142	269
ERS003578	6C	2902	2	10	4	18	6	112	14
ERS006733	6C	2902	2	10	4	18	6	112	14
ERS067700	9L	Novel	2	12	8	8	27	28	5
ERS067702	9L	Novel	2	12	8	8	27	28	5
ERS096174	9L	Novel	2	12	8	8	27	28	5
ERS067701	9L	Novel	2	12	8	8	27	28	5
ERS006735	15B	Novel	54	9	4	5	36	142	269
ERS096155	15B	7522	1	32	97	5	6	88	17
ERS024514	19A	Novel	12	19	2	17	15	22	14
ERS024528	19A	Novel	7	13	53	6	6	6	8
ERS003570	19A	2062	1	5	53	32	14	20	199
ERS096157	19A	2062	1	5	53	32	14	20	199
ERS096158	20	5435	2	5	36	12	6	20	269
ERS096159	20	Novel	13	5	4	5	6	28	168
ERS096162	23B	Novel	33	5	54	63	17	20	168
ERS067685	23B	Novel	2	43	54	5	7	20	6
ERS067684	23B	Novel	2	43	54	5	7	20	6
ERS067686	23B	Novel	2	43	54	5	7	20	6
ERS096176	29	5483	8	11	255	47	6	346	9

The genotype of the carriage variants is compared to wild type invasive serotypes.

The 19A variant (ERS096157) was located on cluster F, together with a well-characterised Malawian 19A invasive serotype (ERS003570), both belonging to sequence type 2062. ST 2062 has previously been isolated from invasive disease in South Africa. The remaining two 19A serotypes existed in different branches on the tree and showed genotypes, which either had no matching alleles with other 19A serotypes (ERS024514) or had only one allele shared with genotype ST 2062 (ERS024528). These results suggest that vaccine serotypes 6B and 19A are genetically very diverse. Pneumococcal vaccination may exert selective pressure on vaccine serotypes such as 6B and 19A, leading to the emergence of novel and non-vaccine serotype, which may compromise the response of these serotypes towards vaccination. The variants of serotype 9L (ERS096174) formed cluster E with the three previously identified 9L Malawian invasive strains. All the 9L were phylogenetically similar and shared the same MLST allelic profile (Table 5.1) with novel sequence type. A query on MLST database (http://www.spneumoniae.mlst.net) generated ST 4914 (9L) and ST 8833 (9L), as the closest match to Malawian 9L. These were isolated in South Africa from blood and cerebrospinal fluid (CSF) of patients diagnosed with bacteraemia and meningitis respectively, and only shared alleles at five of the seven loci with the Malawian 9L serotypes. The 15B variant (ERS096155) did not cluster with any of the invasive serotypes from Malawi, however its sequence type (ST 7522) had one exact match to the carriage 15B/C isolate on MLST database reported in Kathmandu, Nepal.

The two-serotype 20 variants (ERS096159 and ERS096158) appeared on different branches of the tree and had very different allelic profiles. The sequence type for ERS096159 was unknown, however it was closely related to ST 7651 (serotype 20),

a carriage isolate detected in Nepal, and shared alleles at six loci. The variant shared alleles at five loci with ST 5392 (serotype 20) isolated from carriage in Kenya and invasive disease in India. The sequence type for the second serotype 20 variant (ST 5435, ERS096158) has been isolated from invasive disease in South Africa. The 23B variant (ERS096162) was located elsewhere on the tree and not together with 23B invasive serotypes in cluster C, and shared only two of the seven alleles. A search in the MLST database showed close similarity between the Malawian 23B variant (ERS096162) and ST7687 isolated from carriage in Nepal and shared alleles at five loci. The sequence type ST 5483 for the serotype 29 variant (ERS096176) has been isolated from invasive disease in Mozambique. These results show that there is high genotype diversity among serotype variants in carriage and invasive wildtype serotypes in Malawi and that most of the genotypes were novel. A search on the MLST database showed close similarity with Strains from South Africa, Mozambique and Kenya. There is a lot of migration of people amongst these countries, either as migrant workers or traders, which could facilitate the spreading of such strains across these nations.

5.2. Characterising the CPSlocus for serotype 6B variant

Serotype 6B is one of the serotypes covered by the current 13-valent pneumococcal conjugate vaccine, and is the second most common cause of invasive disease globally, after serotype 14 (Johnson et al., 2010). In Malawi, it is second to serotype 1 as the most commonly isolated serotype from sterile sites such as blood and CSF (Cornick et al., 2011). Any changes occurring at the CPS locus of serotype 6B may lead to changes in the structure of the polysaccharide capsule and may render the variant unresponsive to the vaccine-induced immunity (Temime et al., 2008,

Brueggemann et al., 2007a). In this study, carriage 6B variants represented 34% (16/37) of all variant serotypes detected. The 6B variants had previously been isolated from invasive disease in the Netherlands (Elberse et al., 2011), and based on nucleotide differences in genes at the CPS loci, three subtypes of 6B serotypes were defined: 6B-I, 6B-II, and 6B-III (Elberse et al., 2011). To determine whether Malawian 6B variants belonged to any of the above sub-groups, a phylogenetic comparison based only on SNPs within the CPS locus was conducted. The comparison was also extended to previously characterised Malawian invasive 6B strains (Figure 5.2). The phylogeny was divided into three distinct clusters: Cluster 1, Cluster 2, and Cluster 3. The clustering clearly showed that Malawian 6B carriage variants were distinct from the invasive strains. All Malawian invasive 6B strains belonged to cluster 1, while cluster 2 contained invasive 6B variants from the Netherlands. All the strains in cluster 2 were shown to either belong to serotype 6B subtype 1 (6B-I) and/or 6B subtype 2 (6B-II) only (Elberse et al., 2011). The Malawian carriage 6B variants belonged to cluster 3 together with two variants (JF911507.1 and JF911504.1) from the Netherlands, members of 6B subtype 3 (6B-III) (Elberse et al., 2011). This suggested that the Malawian 6B variants belonged to subtype 6B-III. Although all the Malawian 6B variants belonged to 6B-III, they exhibited very distinct genotypes, with virtually no concordance across any of the seven alleles for 6B-III (Table 5.2). These data show that the Malawian 6B variants are phylogenetically distinct from the Malawian wildtype invasive 6B serotypes; however we have shown that the Malawian 6B variants in carriage belong to the serotype 6B subtype III (6B-III) by forming a single cluster with 6B-III from the Netherlands.

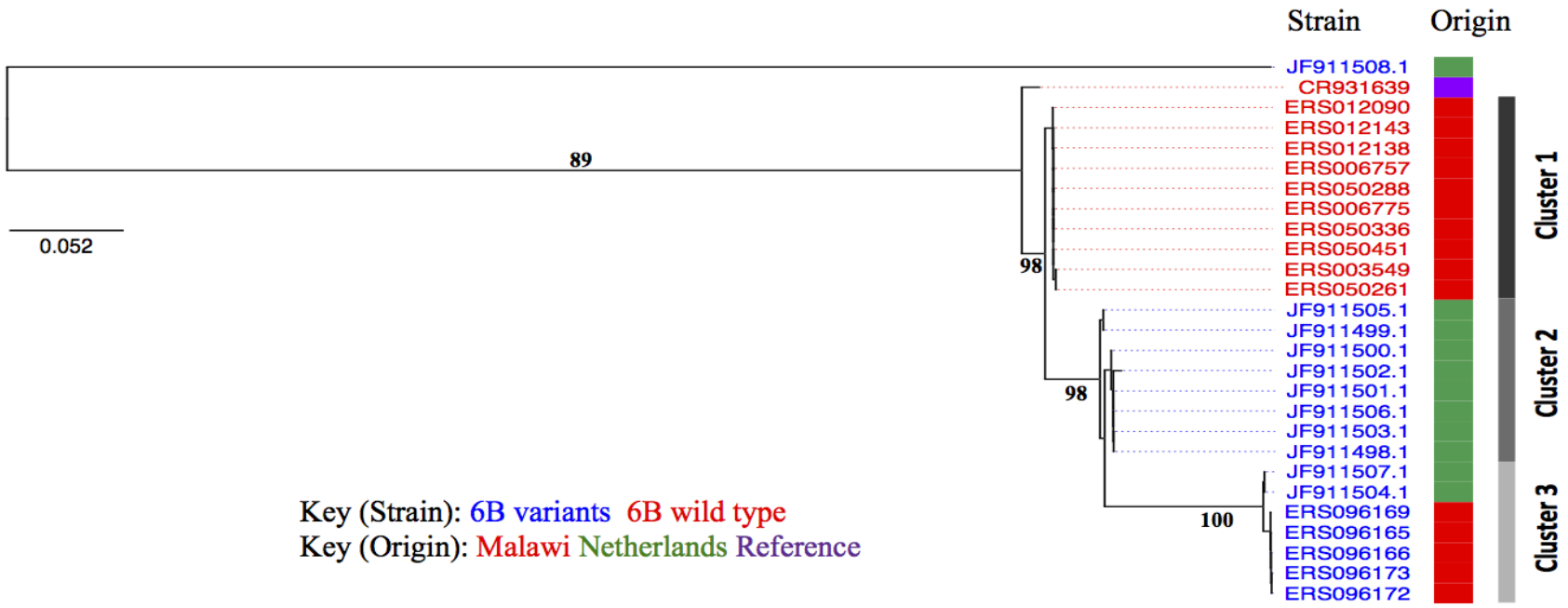


Figure 5.2 Phylogenetic analysis of serotype 6B variants

The tree includes wild type invasive 6B serotypes and carriage 6B variants from Malawi, and invasive 6B variants from the Netherlands. All carriage strains are shown in blue while invasive strains are color-coded red. The origin of the samples is colored red (Malawi) and green (Netherlands). The tree was generated using the alignment of SNPs from the CPSlocus genomic sequence for each serotype with RAxML, by performing 100 bootstraps. The SNPs were identified using reference 6B-III *Streptococcus pneumoniae* strain JF911507 (Elberse et al., 2011).

Table 5.2 MLST profile of the CPS locus variants

Strain ID	Serotype	Country of origin	ST	*aroE*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*	*Closest matching ST**	*Source**	*Region**
JF911504.1	6B-III	Netherlands	Novel	7	6	1	8	6	1	14			Netherlands
JF911507.1	6B-III	Netherlands	90	5	6	1	2	6	3	4			Netherlands
ERS096173	6B	Malawi	Novel	2	8	1	10	17	1	19	2863 (33D) (DLV) 4441 (6C)(DLV) 5266 (22A) (DLV)	Unknown Blood Blood	Denmark Israel Kenya
ERS096172	6B	Malawi	Novel	1	43	29	1	77	1	14	8783 (NT)(DLV)	Carriage	China
ERS096166	6B	Malawi	Novel	10	9	4	5	15	20	28	6382 (6A)(SLV) 7744 (6A)(SLV)	CSF Unknown	South Africa
ERS096165	6B	Malawi	Novel	10	9	4	5	15	4	2	6382 (6A)(DLV) 7744 (6A)(DLV)	CSF Unknown	South Africa
ERS09616	6B	Malawi	Novel	54	5	4	5	36	142	269	8050 (6A)(SLV)	Carriage	The Gambia
ERS096157	19A	Malawi	2062	1	5	53	32	14	20	199		Blood/CSF	South Africa
ERS096159	20	Malawi	Novel	13	5	4	5	6	28	168	5392 (20)(DLV) 7651 (20)(DLV)	Carriage/CSF Carriage	Kenya/India Nepal
ERS096158	20	Malawi	5435	2	5	36	12	6	20	269		Blood	South Africa

The analysis includes two 6B-III subtype references from the Netherlands. Parameters marked (*) show the closest matching ST from the MLST database (http://www.spneumoniae.mlst.net), which include sample source (blood, carriage or csf) as well as country of origin (region).

5.2.1. Genetic recombination within the CPS locus of 6B variants

With reference to the published 6B strain CR931639, all the 6B strains from carriage and invasive disease in Malawi harboured SNPS within their CPS loci; however, the average SNPs density for variants in carriage (50.7 snps/kb/strain) was higher than invasive 6B strains (9.0 snps/kb/strain). Further analysis was conducted to determine if such SNPs translated to genetic recombination and whether differences existed between the variants and invasive 6B strains. Results showed four regions of recombination in all the 6B variants. These occurred in *wzg, wciO, rmlA* and *rmlC* genes (Figure 5.3), which respectively, encode for an integral membrane regulatory protein, putative ribitol transferase, glucose-1-phosphate thymidylyltransferase, and dTDP-4-keto-6-deoxy-D-glucose 3,5-epimerase. All the invasive 6B serotypes showed recombination in *rmlA* gene. Two strains (ERS050261 and ERS003549) showed recombination in *aliB*, tnp and *wzg* genes, while one isolate (ERS012090) showed recombination in *rmlD* gene in addition to the *rmlA* gene. The *aliB* gene is a pseudogene for oligopeptide binding protein. The *tnp* is a transposon, which is a mobile genetic element implicated in horizontal gene transfer, while *rmlD* gene encodes for dTDP-4-keto-L-rhamnose reductase, which is essential for the synthesis of rhamnose sugar, a component of the polysaccharide capsule. All the recombination events in the 6B variants were ancestral, suggesting they have been maintained throughout the subsequent generations. This would suggest they might offer a selective advantage to the variant, thereby enabling it survive its environment and replicate successfully, or they could be neutral with no selective pressure to remove them.

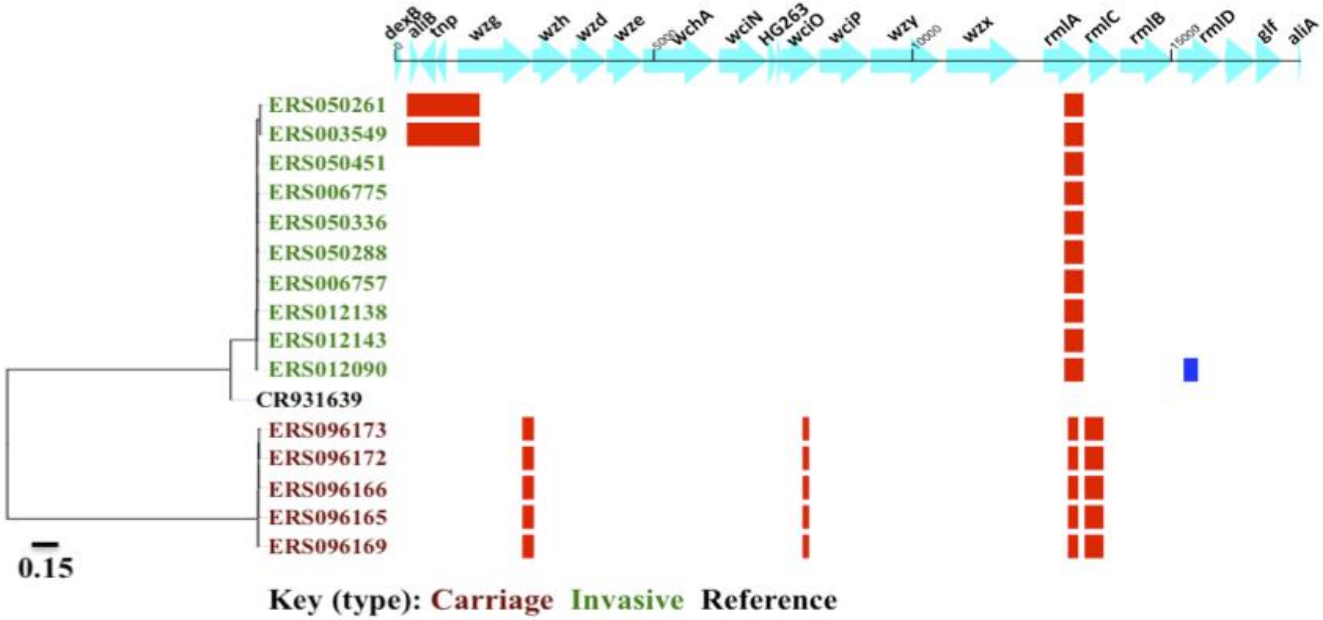


Figure 5.3 Genetic recombination within the CPSlocus of serotype 6B variants

The blocks in red and blue colour indicate regions of recombination within the CPS locus for each isolate. The blue colour block indicates that recombination in that particular region(s) was only detected in one isolate, and therefore a recent event. The red colour blocks shows recombination occurred in more than one strain and therefore an ancestral event. There were more events of recombination observed in carriage CPS locus variants of serotype 6B than invasive strains.

5.2.2. A deletion within *HG263* gene of serotype 6B

The *HG263* gene is a member of the homology group 263 and encodes a putative LicD-family phosphotransferase protein (Bentley et al., 2006). This protein is involved in the transfer of a glycosyl group from activated donor molecule to specific acceptor molecule (http://www.uniprot.org), synonymous with polysaccharide capsule biosynthesis transferase required for repeat unit construction. In this study, the *CPS* locus gene sequences for the reference and variant 6Bs were compared. The sequence identity ranged from 90 to 100% with the lowest sequence identity observed in the first six genes, which includes four initiation genes (*wzg, wzh, wzd, wze*) and two-type specific genes (*wchA* and *wciN*) (Figure 5.4).

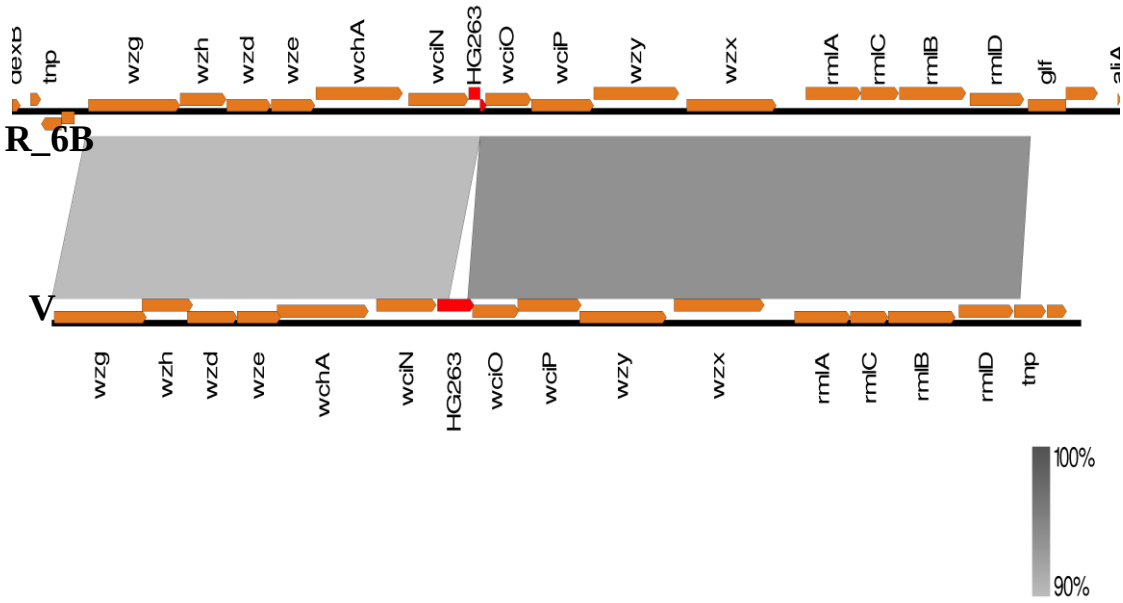


Figure 5.4 The CPS locus genes of serotype 6B variant

A 300 base pair deletion within the *HG263* gene was observed in the wild type 6B (R) (CR931639) and not in the 6B variant (V, ERS096173). The figure was generated by EasyFig® (Sullivan et al., 2011).

A genetic variation was detected in the *HG263* gene between 6B variants and the reference wild type 6B serotype. The carriage 6B variants contained an intact allele of the *HG263* gene, which was about 579 base pairs in length. The reference and all invasive 6B serotypes harboured a HG263 gene with a 300 base pair deletion resulting in a frame shift mutation (Figure 5.4). This gene was annotated as a *pseudogene* and therefore non-functional (Bentley et al., 2006). Mutants of LicD-family phosphotransferase gene have been reported with a reduced ability to take up choline; decreased ability to adhere to host cells and to be less virulent (Zhang et al., 1999). Thus by harbouring an intact allele of the *LicD* gene, it is not clear whether the variants will have an enhanced ability to colonise the nasopharynx or cause invasive disease. To test the validity of this observation, human mouse model experiments were conducted using the wild type 6B strain BHN418 (ST138) as control.

5.3. Testing the invasive ability of Malawian serotype 6B variants

The ability of pneumococcal 6B variants to cause invasive disease was tested in mouse models (Chapter 2, section 2.9), using the carriage 6B (BH418) strain as reference. For comparison, previously characterised invasive 6B serotypes from Malawi were also tested. Results showed that the 2 invasive 6B serotypes (ERS012143/ERS050336) were the most able to cause lethal systemic disease after intranasal challenge and caused death of all mice after 50 hours (Figure 5.5).

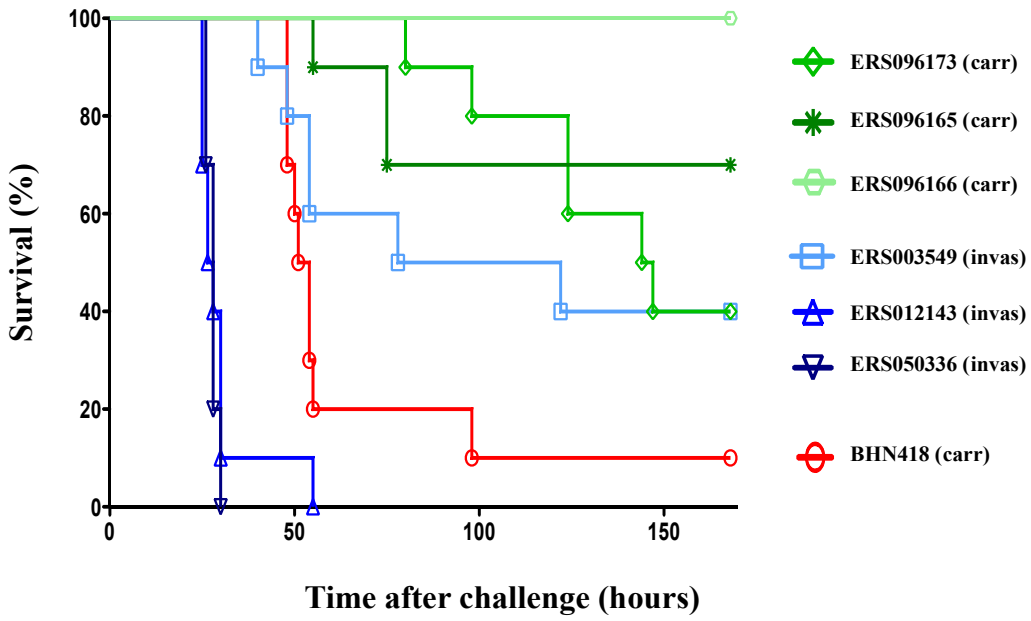


Figure 5.5 Invasive ability of serotype 6B variants

The Kaplan-Meier plot showing the invasive ability of serotype 6B variants. Mice (n=10 per bacterial isolate) were infected intranasally with 10^6 CFU *S. pneumoniae* per mouse in a volume of 50µl. The graphs in green represent mice challenged with three different carriage 6B variants strains, whilst graphs in blue represent mice inoculated with three different 6B invasive strains. The graph color-coded red represents the reference carriage 6B sample. Mice survival was monitored up to 168 hours, at which point the experiment was ended.

This was followed by control 6B serotype (BH418, ST138), which caused death of 90% of mice after 150 hours. The invasive strain ERS096173 caused death to 60% of the mice challenged and was the least able to cause invasive disease among the three invasive strains tested. All the mice challenged with the carriage 6B variant strain ERS096166 survived throughout the entire assay period, suggesting inability by this strain to cause disease. The remaining 2 carriage strains: ERS096165 and

ERS096173 were able to cause systemic disease after intranasal challenge; however the latter was most lethal as it caused death to 60% while the former only managed to kill 30% of the mice understudy. The mortality rate by the carriage 6B strain ERS096173 was similar to that for invasive 6B strain ERS003549 after 150 hours, however death for mice challenged with the invasive strain occurred much earlier. Analysis by log-rank (Mantel-Cox) test showed that survival curves were significantly different ($p<0.0001$). Further, a comparison of the Kaplan-Meier curves showed a significant trend in survival scores across the carriage 6B variants and wild type invasive 6B strains ($p<0.0001$). Thus overall, the survival rate of mice challenged with the 6B variants was higher than the rate for the invasive 6B strains, suggesting the variants were less invasive and less able to cause disease compared to invasive 6B types.

5.4. An inversion of *rmlD* gene in serotype 19A variant

Serotype 19A is also one of the PCV-13 serotypes and globally it is ranked number 8 as the most invasive disease causing serotype (Johnson et al., 2010). Serotype 19A was a major cause of serotype replacement disease in various settings post PCV-7 introduction (Cohen et al., 2011, Brueggemann et al., 2007a). Microarray serotyping detected serotype 19 as a CPS locus variant with missing *rmlD* gene. The *rmlD* gene together with three other similar genes (*rmlA, rmlB, rmlC*) encodes for proteins involved in the synthesis of rhamnose sugar, a component of the pneumococcal capsule (Graninger et al., 1999). The molecular structure of wild type serotype 19A polysaccharide capsule is composed of rhamnose, N-acetylmannosamine, and glucose, as a repeat unit (Bentley et al., 2006). The pattern of SNPs within the 19A CPS locus variant and *rmlD* gene inversion was identical to serotype 19A-I subtype

(Elberse et al., 2011), however the sequence types were different. To investigate genetic changes within the 19A CPS locus, whole genome sequences were obtained and the CPSgenes were annotated in artemis (Rutherford et al., 2000). The genetic comparison diagrams were then generated using EasyFig (Sullivan et al., 2011). Results showed ~100 % sequence similarities in the CPS genes for the 19A variant and 19A reference, except for *rmlD* gene which showed only 77% sequence similarity (Figure 5.6). The lower sequence similarity signifies mismatches in the sequences possibly due to SNPs within the gene or it could be attributed to the fact that the *rmlD* gene in the variant was oriented in the opposite direction compared to other *rml* genes and the rest of the CPS locus genes.

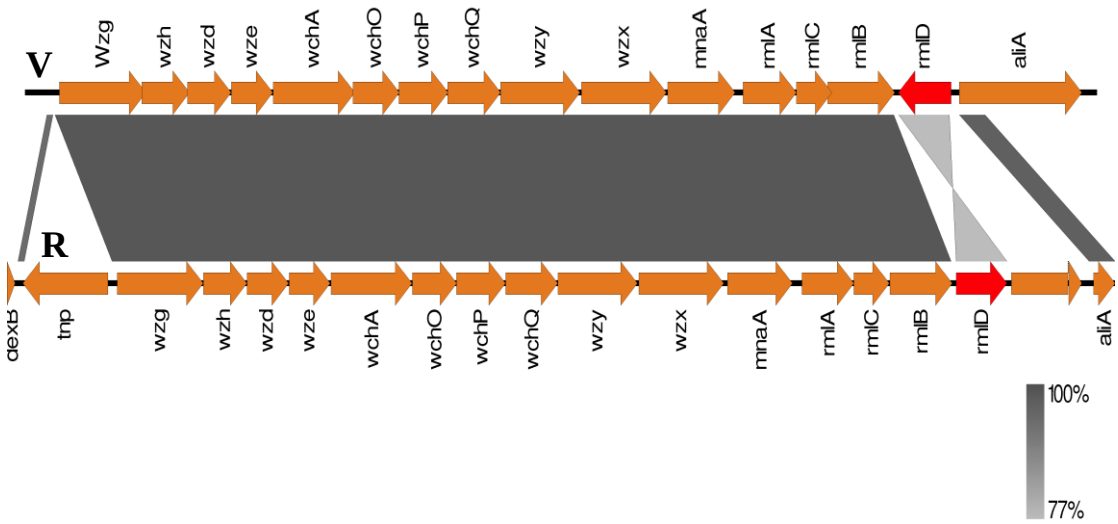


Figure 5.6 The CPS locus genes of serotype 19A variant

The CPS *locus* genes of the 19A variant (V, ERS096157) compared with a reference serotype 19A (R) CR931675). An inversion in *rmlD* gene (red colour) was observed in the variant. The figure was generated using EasyFig® (Sullivan et al., 2011).

5.5. A deletion within the *whaF* gene in serotype 20 variant

In this study, two serotypes belonging to sero-group 20 were detected as CPS locus variants by microarray serotyping. The CPS gene sequence of the two variants of serotype 20 showed a 716 base pair deletion within the *whaF* gene, and their sequences identity with the reference ranged from 76 to 100% (Figure 5.7). The *whaF* gene is located between the *wciD* and *wzx* genes within the CPS locus and is 963 base pairs in length. It encodes for putative glycosyl transferase (GT) protein (Bentley et al., 2006), which mediates the transfer of glycosyl (αGlc) residue in the capsular polysaccharide repeat unit. The serotype 20A subtype contains a mutation in the *whaF* gene which leads to loss in gene function (Calix et al., 2012b). Through a deletion of the *whaF* gene, the variant in our dataset would be expected to have a capsular polysaccharide structure similar to 20A subtype. However, further analysis would be required to elucidate the actual capsular polysaccharide structure, as well as their ability to colonise or cause invasive diseases.

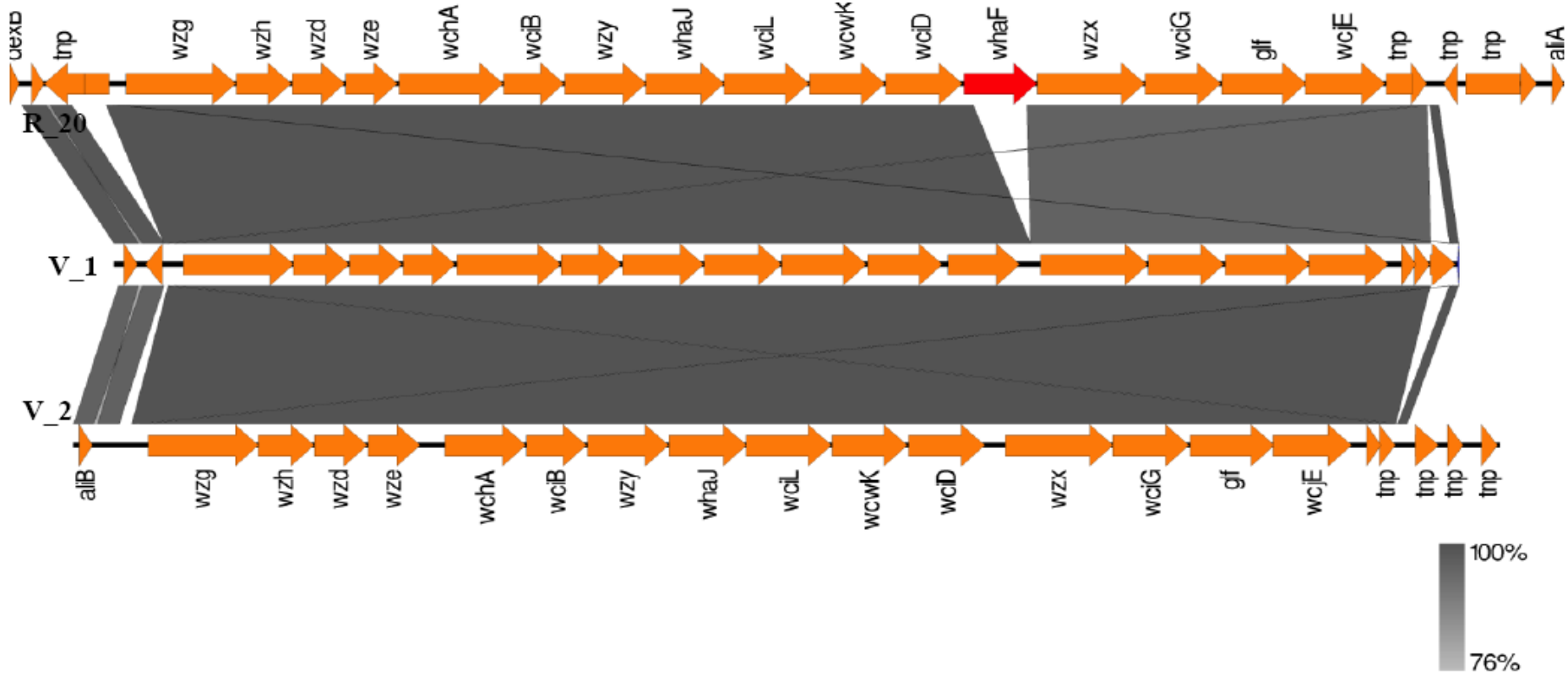


Figure 5.7 The CPS locus genes of serotype 20 variant

The CPS locus genes of serotype 20 variants from a child (V_1, ERS096159) and adult (V_2, ERS096158) was compared to reference serotype 20 (R) (CR931679). A 716 base pair deletion of the *whaF* gene (color-coded red) was observed in the variants. The figure was generated using EasyFig® (Sullivan et al., 2011).

5.6. Novel serotype containing genes from serotype 29 and 35B

The pneumococcus is a highly transformable organism that readily undergoes genetic recombination through horizontal gene transfer. Multiple carriage is thought to be a pre-requisite for horizontal gene transfer. Evidence of genetic recombination in the pneumococcus has been widely reported (Croucher et al., 2013, Everett et al., 2012, Hanage et al., 2009, Marks et al., 2012). Recombination involving antibiotic resistance and CPS locus genes has a bearing on treatment strategies. Novel serotypes containing a mosaic cluster of CPS genes from different serotypes have been reported (Salter et al., 2012). In this study, a novel variant serotype containing all the 12 CPS genes from serotype 29 and a *wciG* gene from 35B was obtained (Figure 5.8, V). The novel serotype showed a higher sequence similarity to serotype 29 than 35B and it is likely that the *wciG* gene occurred as an insertion into the CPS locus of serotype 29. The *wciG* gene encode for *O*-acetyltransferase enzyme, which is responsible for *O*-acetate substitution on the galactofuranose (Gal*f*) residues in the repeat unit of the polysaccharide capsule (Calix et al., 2012b). The repeat unit of the capsular polysaccharide for the wild type serotype 29 consists of three molecules of Gal*f* and one molecule of *N*-acetylgalactosamine pyranose (GalpNAc) without acetylation (Bentley et al., 2006). Therefore in our setting, there is a possibility that the incorporation of *wciG* gene into the CPS locus of serotype 29 would introduce *O*-acetylation at the Gal*f* moiety of the capsular polysaccharide. Consequently, this may enhance the ability of the 29 variant to successfully colonise the nasopharynx, as it will be expected to evade the host lysozyme activity.

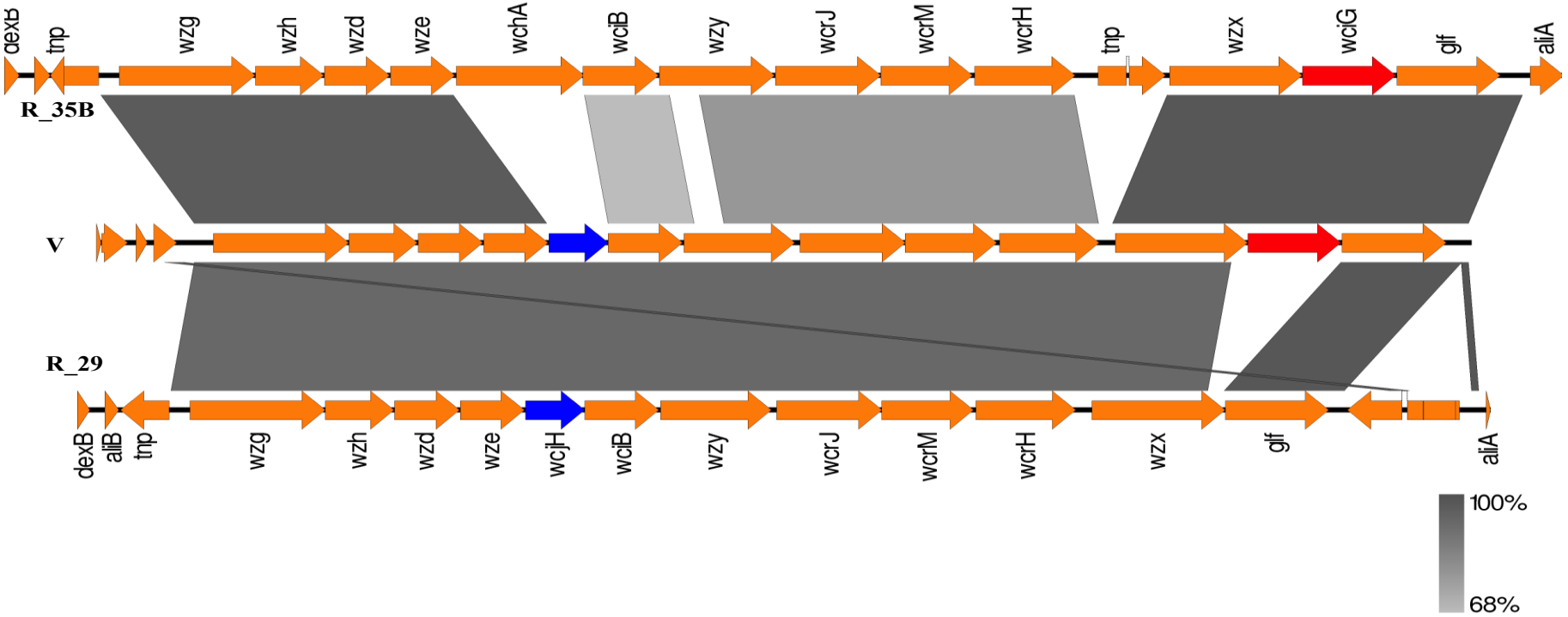


Figure 5.8 The CPS locus genes of serotype 29 variant

The CPS locus genes of serotype 29 variant (V, ERS096176) were compared with reference types 29 (R_29) (CR931694) and 35B (R_35B) (CR931705). The variant from an HIV positive adult single carrier, showed a possible serotype 29 and 35B hybrid by harbouring both *wcjH* (blue), unique to serotype 29 and *wciG* (red), unique to serotype 35B. The figure was generated using EasyFig® (Sullivan et al., 2011).

5.7. Characterisation of CPS locus variants of serotypes 9L, 15B and 23B

Serotypes 9L, 15B and 23B are not included in the current vaccine; however, they have been reported to cause invasive disease elsewhere (Rodgers et al., 2009). The microarray serotyping reported divergence in *wcjE*, *wzy*, and *wcjB* genes in 9L variant. However, analysis of whole genomes of serotype 9L only showed a mismatch in *wzy* gene, and the sequence similarity ranged from 83-100% for the entire CPS locus (Figure 5.9). The *wzy* gene encode for polymerase enzyme, which is responsible for linking individual repeat units to form lipid-linked polysaccharide capsule (Bentley et al., 2006). Any variation in this important gene is likely to compromise the synthesis of the polysaccharide capsule, which could render the variant un-encapsulated. Serotype 15B was determined by microarray to be divergent in *wchA* and *wchK* genes. Unexpectedly, analysis of CPS locus sequences showed high sequence similarity (80-100%) between the 15B variant and the reference (Figure 5.10). The anomaly detected by microarray could be attributed to cross-hybridisation, which is a common occurrence in assays involving microarray. The variant of serotypes 23B were suspected to have genetic variations in gene *wchX*, *rmlA* and *rmlC* gene. Sequence comparison showed a mismatch in *wchW* gene only, with percentage sequence identity ranging from 84% to 100% (Figure 5.11). The *wchW* gene encodes for a putative glycosyltranferase protein, which also belongs to type-specific genes. Changes in this gene may impair capsule synthesis or lead to the synthesis of a novel capsule, consequently affecting the pneumococcal phenotype.

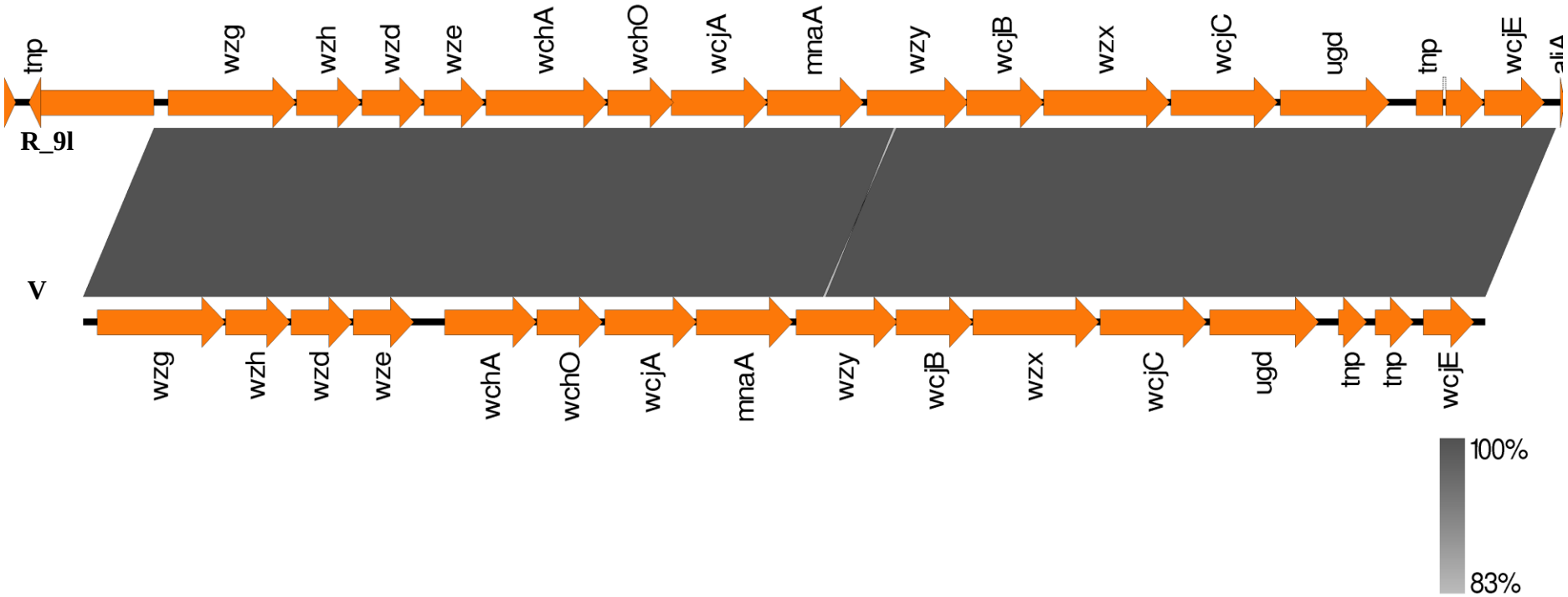


Figure 5.9 The CPSlocus genes of serotype 9l variant

The CPS locus genes of serotype 9l variant (V, ERS096174) compared with a reference serotype 9l (R_9l) CR931646). The figure was generated using EasyFig® (Sullivan et al., 2011).

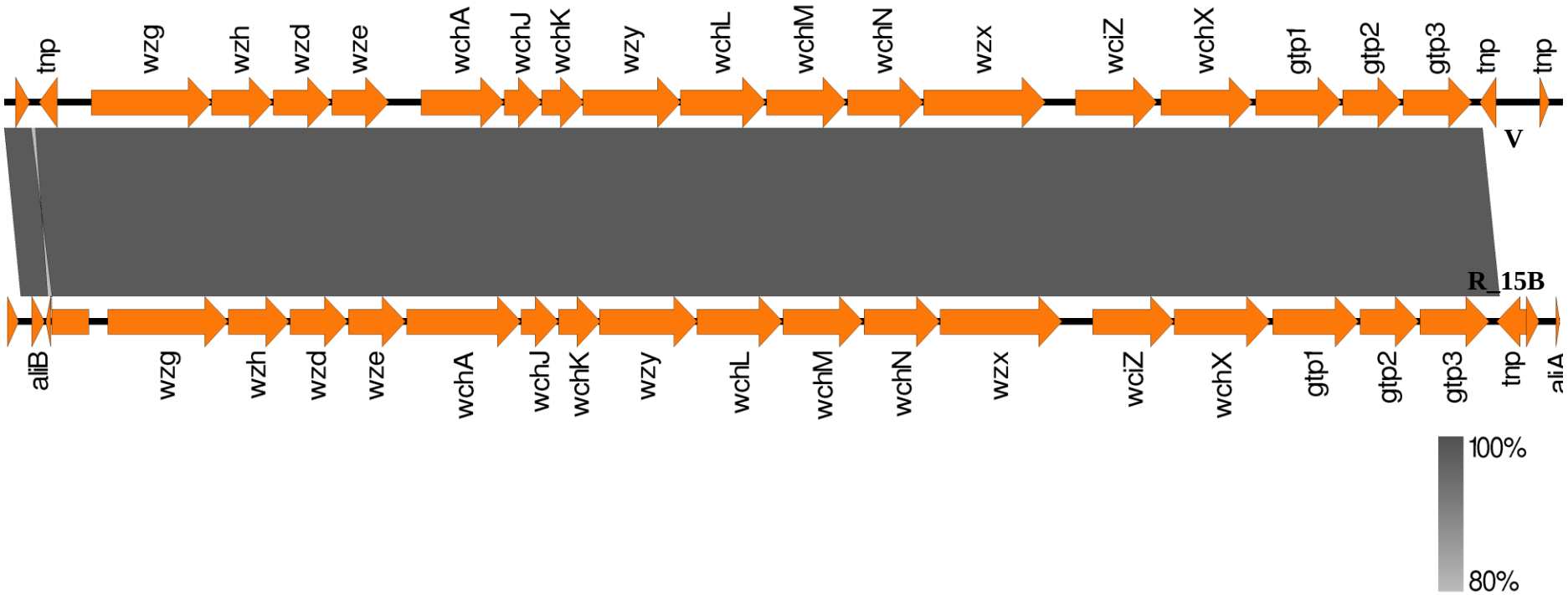


Figure 5.10 The CPS locus genes of serotype 15B variant

The CPS locus genes of serotype 15B variant (V, ERS096155) compared with a reference serotype 15B (R_15B) CR931664). The figure was generated using EasyFig® (Sullivan et al., 2011).

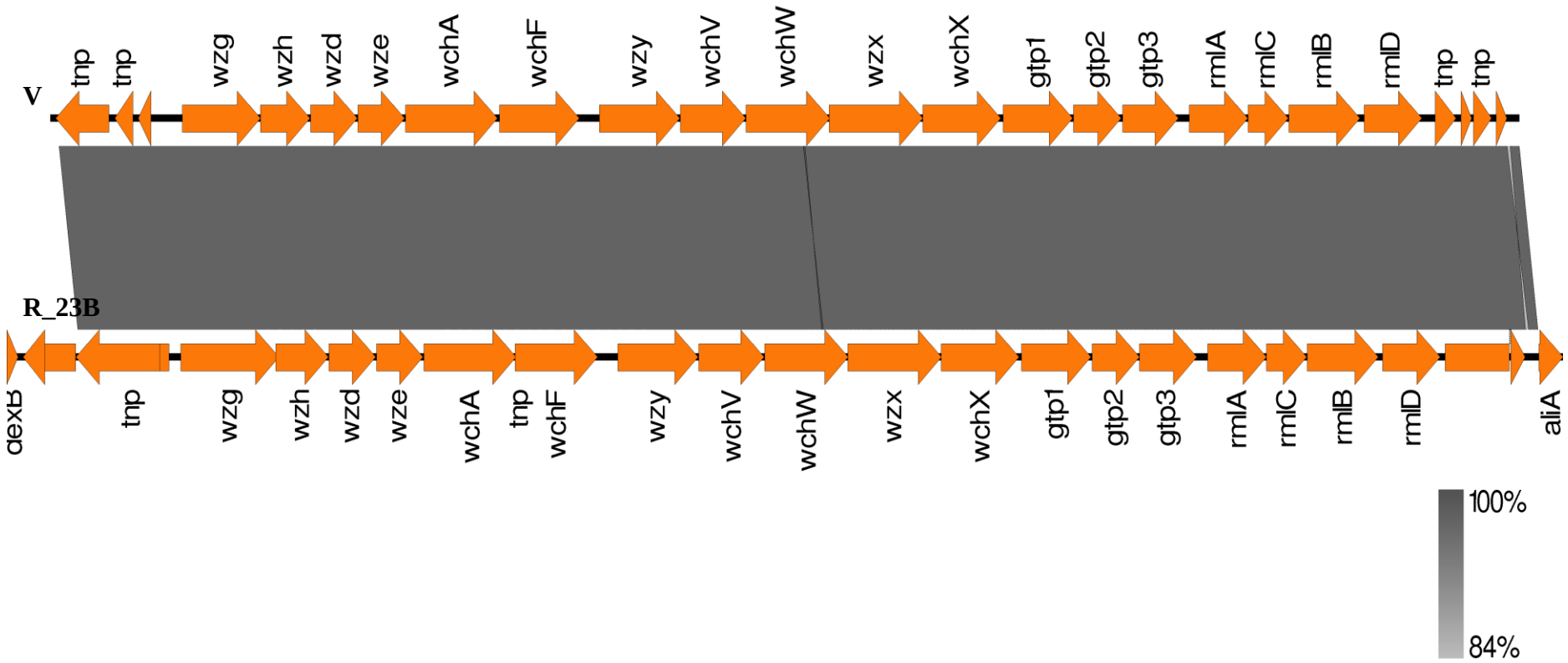


Figure 5.11 The CPS locus of serotype 23B variant

The CPS locus genes of serotype 23B variant (V, ERS096162) compared with a reference serotype 23B (R_23B) CR931684). The figure was generated using EasyFig® (Sullivan et al., 2011).

5.8. Discussion

The pneumococcal polysaccharide capsule defines the serotype (Bentley et al., 2006) and is essential for the successful colonisation and transmission of *S. pneumoniae* by acting as a barrier to phagocytic killing by the host immune system (Weinberger et al., 2009). The capsule is also a target for current conjugate vaccine formulations. To survive under such conditions, the pneumococcus can alter its CPS locus, through genetic recombination, which could ultimately lead to vaccine escape through serotype switching. In Malawi, unrelated serotypes with the same sequence type have been isolated from invasive disease in Malawi, suggesting occurrence of possible recombination within the CPS locus pre-vaccination (Everett et al., 2012). In the present study, the CPS locus variants of vaccine-associated serotypes 6B, 19A and 20, and non-vaccine serotypes 29, 9l, 15B and 23B were detected. Serotype 6B is among the top five serotypes in circulation together with serotypes 1, 5, 6A and 14, cumulatively responsible for over 50% of all invasive disease in Africa, Asia and Europe (Johnson et al., 2010). In Malawi, serotype 6B is the second most isolated strain from invasive disease in children after serotype 1 (Cornick et al., 2011). In our data set, serotype 6B variants were genetically distinct from invasive 6B serotypes and had a significantly high SNPs density within the CPS locus. There were also more recombination events occurring in carriage 6B variants than the wild-type invasive strains. Genetic recombination involves the integration of foreign DNA in the chromosome of a recipient cell, mediated by bacteriophages (transduction) (Canchaya et al., 2003), direct cell-to-cell contact (conjugation) (Chen et al., 2005), or uptake of exogenous DNA from the environment (transformation) (Claverys et al., 2009), and it is reported to occur primarily in carriage. Such levels of recombination in 6B could be driven by the high multiple carriage rates observed in the Malawian

setting (Donkor et al., 2011). Of particular concern is that such recombination is occurring in the absence of vaccine selection pressure, and may naturally lead to the generation of vaccine escape strains, thereby negatively affecting the efficacy of the vaccine. In mouse models, the 6B variants are showing clear indications of being less invasive. In addition, each of the 6B variants exhibited the sequence types not reported elsewhere, which makes their invasive potential unknown. One key difference in genes within the CPS locus was the presence of an intact allele of the *HG263* gene in all the 6B variants, while the wild type invasive 6B strains and the reference contained a base pair deletion (~300bp) within the *HG263* gene resulting in a frame shift mutation (Figure 5.4). *HG263* gene is a member of the LicD transferase family of proteins involved in lipopolysaccharide biosynthesis (Lysenko et al., 2000). LicD mutants have a reduced ability to take up choline; decreased adherence to host cells and are less virulent (Zhang et al., 1999). Thus by harboring an intact *LicD* gene, our 6B variants may have a colonisation advantage over the wild type, which may facilitate increased invasive potential and/or allow for vaccine escape. However, in the absence of other factors such as vaccination, antibiotic use and HIV infection, the trends in colonisation and invasive potential may decrease due to the acquisition of nonspecific immunity to the pneumococcus by the host (Cobey and Lipsitch, 2012). The potential of the 6B variants to cause invasive disease was investigated in mice, through intranasal challenge (Figure 5.5). Whereas most mice infected with invasive 6B serotypes died, the survival rate among mice challenged with 6B variants was very high suggesting that the variants were less invasive when compared to either the wild type invasive strains or the reference carriage BHN418 6B strain. Having established that the variants were poor at causing invasive disease, further work is required to investigate the ability of the variants to colonise the

nasopharynx through monostrain carriage experiments. In addition, further investigations using mouse models are still required to establish how the variants will respond towards vaccination, through intranasal challenge of unvaccinated mice and those vaccinated with PCV13.

The CPS locus variant of serotype 19A was shown to have an inversion in the *rmlD* gene (Figure 5.6). The *rmlD* gene is part of a four CPS gene cluster including *rmlA*, *rmlB* and *rmlC*, which encode four enzymes required in the synthesis of dTDP-L-rhamnose from D-glucose 1-phosphate (Li et al., 2003). Although inversions are known to maintain the overall amount of the genetic material (Griffiths et al., 2000), recent findings suggest gene inversions may lower the expression level of the affected gene, resulting in abnormalities at phenotypic level (Naseeb and Delneri, 2012). An inversion in *rmlD* gene for serotype 19A has been reported previously (Elberse et al., 2011, Morona et al., 1999), with the promoter sequence located upstream of the *rmlD* gene (Morona et al., 1999). Other streptococcal species such as *mitis* and *oralis* have also been shown to contain an *rmlD* gene, which is oriented in the opposite direction (Yoshida et al., 2005). In our setting, the *rmlD* gene inversion in the 19A variant may impair the function of the whole *rml* gene cluster, necessary for rhamnose biosynthesis. Failure to synthesise rhamnose may render the 19A variant non-capsulate, since rhamnose forms part of the repeat unit structure for the polysaccharide capsule (Bentley et al., 2006). This could compromise the response of the 19A variant towards vaccine-induced immunity.

Recently, a capsular polysaccharide structural difference has been reported within serogroup 20, due to truncation and loss of function of the *whaF* gene (Calix et al.,

2012b). The *whaF* gene is located between the *wciD* and *wzx* genes (Figure 5.7), and encodes for putative glycosyl transferase (GT) protein (Bentley et al., 2006). The truncation in the *whaF* gene correlated with the loss of an αGlc residue in the capsular polysaccharide repeat unit of serotype 20A (Calix et al., 2012b). In contrast, the Malawian serotype 20 variants had a 716bp deletion within the *whaF* gene. This deletion would render the *whaF* gene non-functional, leading to a loss of an aGlc residue in the capsular polysaccharide repeat unit of the variant. This suggests the serotype 20 variants circulating in Malawi belong to subtype 20A. It is unclear how this deletion will impact on the invasiveness of this serotype 20 variant, although studies have shown that invasive strains of serotype 20B harbour an intact allele of the *whaF* gene (Calix et al., 2012b).

The presence of a novel serotype in our dataset, which contained CPSgenes from serotype 29 and a single *wciG* gene from serotype 35B (Figure 5.8) provided evidence of interspecies horizontal gene transfer occurring naturally in Malawi. The *wciG* gene is responsible for the acetylation of the galactofuranose (Gal*f*) residues in the repeat unit of the pneumococcal polysaccharide capsule. Acetylation describes a reaction that introduces an acetyl functional group into a chemical compound and has been reported to occur in 90% and 80% of bacteria and human proteins respectively. Acetylation can be used to modulate gene functionality, and also occurs as a co-translational and post-translational modification of proteins. The cell of several bacterial species including *S. pneumoniae* is surrounded by a peptidoglycan exoskeleton, which consists of a polymer of the dissacharide N-acetylmuramic acid $(MurNAc)^2$-(*β*-1,4)- N-acetylglucosamine (GlcNAc) and associated stem peptides (Laaberki et al., 2011). This polymer is responsible for maintaining the shape of the

bacteria and its resistance to cytoplasmic turgor pressure. Peptidoglycan has been a target of many antimicrobial compounds including lysozyme, which is present in mucosal secretions and acts as the major effector of innate immunity. To inhibit the hydrolytic activity of lysozyme, bacteria use several mechanisms including *O*-acetylation of the MurNac moiety of peptidoglycan (Laaberki et al., 2011). A direct correlation between the levels of *O*-acetylation and resistance to lysozyme activity has been reported in several bacterial species such as *Staphylococcus aureus* (Bera et al., 2006) and *Neisseria gonorrhoea* (Rosenthal et al., 1983). By acquiring a *wciG* gene, the polysaccharide capsule of serotype 29 variant would be acetylated at the Gal*f* moiety. This may consequently enhance the ability of the 29 variant to successfully colonise the nasopharynx, as it will be expected to evade the lysozyme activity by the host. Whether the additional gene will alter the invasiveness of serotype 29 variant is subject to further investigation.

5.9. Conclusion

This chapter has highlighted the existence of high levels of genetic diversity in strains belonging to the same serotype in Malawi, as indicated by their phylogenetic clustering as well as multi-locus sequence typing. Also highlighted is the occurrence of genetic recombination within the CPS locus, which could have arisen through horizontal gene transfer due to high levels of multiple carriage in Malawi. Most importantly, such genetic variations occurred in the absence of vaccine selection pressure. For a resource poor setting such as Malawi, genetic changes at the CPS locus can be disastrous as they have the potential to generate strains that may be unresponsive to vaccine treatment strategies. Recently, Malawi enrolled PCV13 to under-five children and naturally, the pneumococcal strains would be expected to

alter their genetic make-up as a survival mechanism, and it would be important to see how the CPS loci of vaccine serotypes change once the vaccine takes full effect. Although this study has revealed the occurrence of genetic variations within the CPS*loci* of various serotypes in the absence of vaccine selection pressure, it has not been able to establish the full extent of the prevalence of the variants due to limited sample size. In-addition, this study lacked sequenced wild type serotypes isolated from carriage in Malawi, which would be an ideal reference for comparing variations at the CPS locus and other characteristics. To address the listed limitations, work is currently underway to characterise serotype variants on a larger scale and also determine their ability to colonise and cause invasive diseases using mouse models.

Chapter 6: Nasopharyngeal microbiota in Malawian children and adults

6.0 Introduction

The human body is colonised by a vast array of microorganisms, which are largely commensal. As discussed in the introduction (chapter 1, Section 1.8), these play important roles in the development of the host immune system and aid metabolism (Morgan and Huttenhower, 2012). The role of microbiota in human health and disease is currently not well documented, however, studies have shown a shift in the composition of the gut microbiome under disease conditions such as Kwashiorkor (Smith et al., 2013) and inflammatory bowel disease (Walker et al., 2011). A similar shift has also been reported in the nasopharyngeal microbiota during acute otitis media (Hilty et al., 2012). Seasonality has also been implicated in driving changes in microbial composition (Bogaert et al., 2011). The nasopharynx is also colonised by a wide range of microorganisms (Kwambana et al., 2011, Pettigrew et al., 2008). However, previous studies in both adults and children from different settings have consistently reported carriage in the upper respiratory tract of at least one of the following five phyla: *Firmicutes*, *Proteobacteria*, *Bacteroidetes*, *Actinobacteria* and *Fusobacteria* (Bogaert et al., 2011, Lemon et al., 2010, Frank et al., 2010, Gong et al., 2013). The major differences in various settings existed in the relative proportions of each phyla which could be attributed to a number of different factors that include environmental conditions such as seasonality (Bogaert et al., 2011, Harrison et al., 1999), methodological factors that range from sampling to analysis (Rapola et al., 1997) as well as upper respiratory tract infections (Syrjanen et al., 2001). The basic characteristics of each of the aforementioned phyla are outlined

below based on information from NCBI (http://www.ncbi.nlm.nih.gov): *Firmicutes* are Gram-positive bacteria consisting of three major classes: *Bacilli, Clostridia* and *Erysopelotrichi*. *Proteobacteria* are considered a major group of bacteria and they are all Gram-negative. The six major classes that constitute *Proteobacteria* are *Alphaproteobacteria, Betaproteobacteria, Acidithiobacillia Gammaproteobacteria, Deltaproteobacteria, and Epsilonproteobacteria. Bacteroidetes* are Gram-negative bacteria and they consist of three large groups: *Bacteroidetes, Flavobacteria* and *Sphingobacteria*. *Actinobacteria* are Gram-positive organisms, which are grouped into one major class *Actinobacteria*; while *Fusobacteria* are Gram-negative bacilli also grouped into one class *Fusobacteria*. In the study presented in this thesis, we investigated for the first time the profiles of the nasopharyngeal microbiota in Malawian children and adults, in a setting with high prevalence of HIV infection and invasive pneumococcal disease. The results showed that four phyla *Firmicutes, Proteobacteria, Bacteroidetes and Actinobacteria* dominated nasopharyngeal microbiota in Malawi.

6.1. Methods

The detailed methods and baseline characteristics of the subjects recruited (Table 6.1) are as previously outlined (Chapter 2, Sections 2.1 and 2.2). Briefly, samples from Karonga were collected from infants in their communities, while samples in Blantyre were collected from children and adults reporting to the Queen Elizabeth Central Hospital as outpatients. Only carriers of *S. pneumoniae* were included in the analysis, which may have a fundamental impact on carriage dynamics of other microbiota in the nasopharynx.

Table 6.1 Baseline characteristics of children and adults sampled

	Location	HIV Status		Total (n)
		HIV-	HIV+	
Children	Karonga	16		16
	Blantyre	10	25	35
Adults	Blantyre	19	36	55
	Total (n)	45	61	115

The subjects included HIV positive adults and children who were also on cotrimoxazole prophylaxis. The use of cotrimoxazole prophylaxis could potentially alter the profile of microbiota in carriage (Varon et al., 2000). The DNA was extracted directly from the nasopharyngeal swabs taken from both adults and children (Chapter 2, section 2.3); followed by a PCR amplification of the V3-V5 variable region of the 16S rRNA gene using barcoded reverse primers (Appendix 4). The 454 sequencing platform was employed to sequence the 16S rRNA PCR amplicons. The sequences generated were analysed using Mothur (Schloss et al., 2009) (Appendix 5) to remove the sequencing artefacts, align and cluster the sequences into operational taxonomic units, and describe the alpha diversity of the nasopharyngeal samples.

6.2. Results

6.2.1. Sampling depth

To establish if sampling was adequate (sampling depth) to account for all the phylotypes present in the nasopharynx of adults and children, rarefaction curves were generated (Gotelli and Colwell, 2001). Rarefaction curves are mainly used to compare the number of species between distinct groups of samples with varying number of samples (Lee et al., 2007). These are constructed by plotting the number of species in each group of samples obtained by randomly subsampling the reads from each group at different levels or frequencies (Gotelli and Colwell, 2001). Initially, all the curves increase rapidly with an increase in the sample size and gradually begin to plateau as more and more samples are added. The initial increase is attributed to identification of dominant species; while only rare and very few species are identified with an increase in sampling depth, resulting in graphs reaching a plateau (Colwell et al., 2012). In the present study, rarefaction curves were used to assess the optimum number of sequence reads to account for all operational taxonomic units (OTUs) (Section 6.1.2) in a given sample. Results showed that further increase in the number of sequence reads beyond the 1000 value did not have a significant impact on the number of OTUs obtained (Figure 6.2). The assumption was therefore made that subsampling 1000 16S rRNA sequence reads per sample would account for most OTUs present in both adults and children's samples.

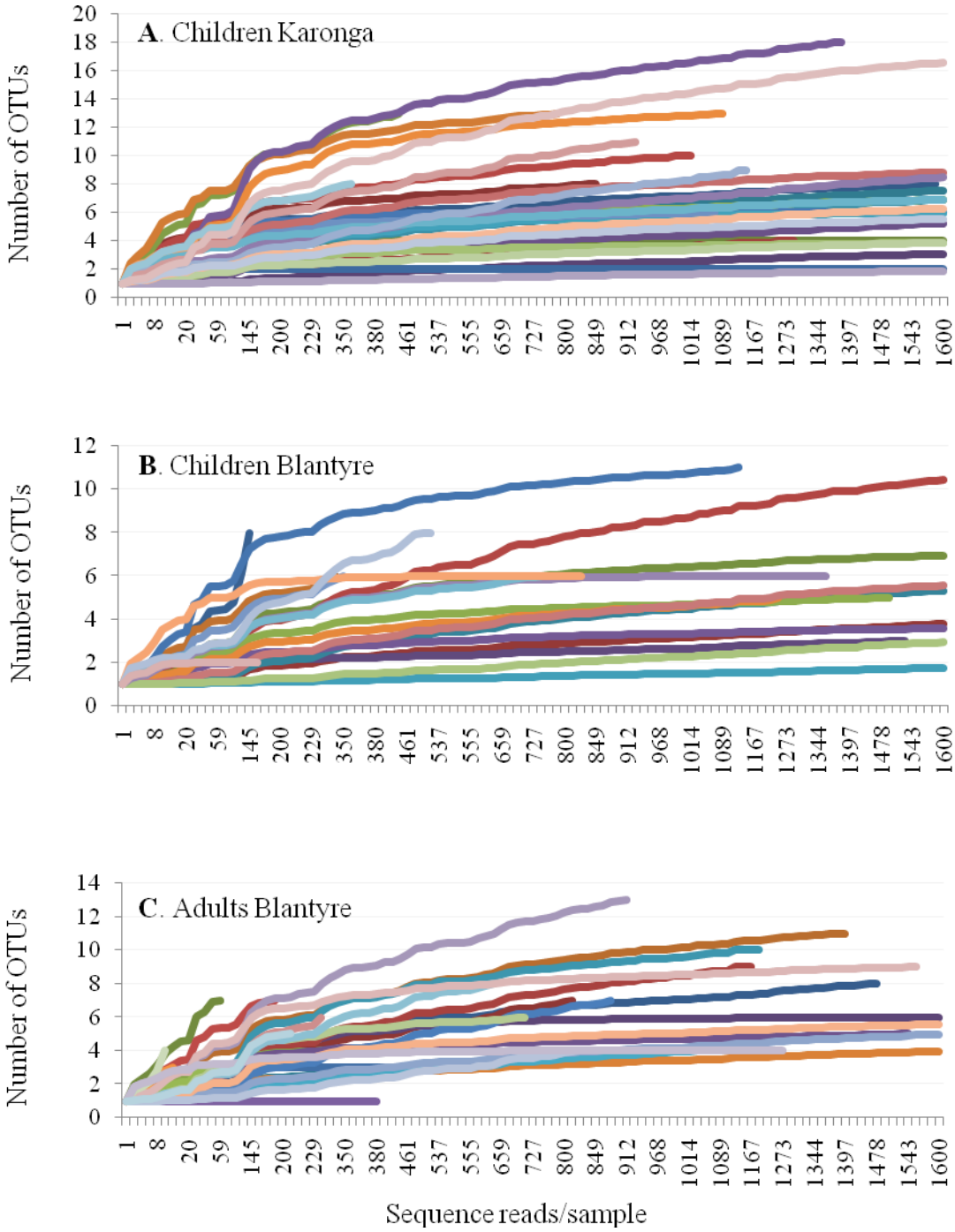


Figure 6.1 Rarefaction curves

Rarefaction curves comparing the number of reads with the number of OTUs (phylotypes) found in the DNA isolated from the nasopharynx of Malawian children (A and B) and adults (C). Each line graph corresponds to an individual sample. Overall, graphs formed a plateau beyond 1000 reads.

6.2.2. Operational taxonomic units

The nasopharyngeal microbiota was characterised by sequencing of the PCR amplified 16S rRNA gene (Chapter 2, section 2.10.6). The 454 sequencing generates sequence tags that are divergent or nearly identical. All the nearly identical sequences, which are thought to represent a taxonomic clade, are placed in one cluster and are collectively known as the Operational Taxonomic Unit (Morgan and Huttenhower, 2012). In this study, the Mothur software package (Schloss et al., 2009) was employed to assign 16S rRNA sequences to OTUs, using the default furthest neighbour approach at a distance of 3%. The total number of 454 sequence reads was 214 754.Of the sample set, 61 were below the ideal read count (≤1000 reads) while 81 were acceptable as their count was above the threshold value of 1000 reads (Figure 6.2).

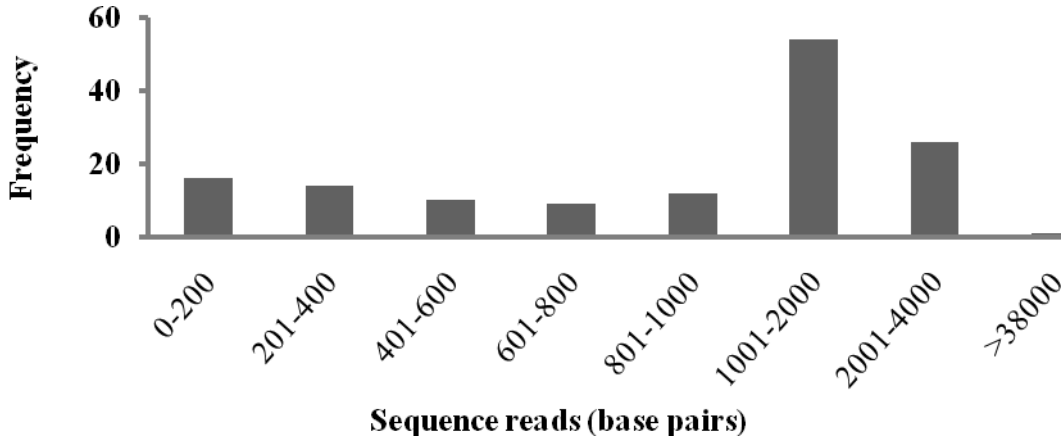


Figure 6.2 Sample frequency versus 454 sequence reads.

Samples with >1000 reads (81 sequences in total) were considered to have an adequate sampling depth for further analysis.

The sequence reads were subsampled to 1000 reads per sample which based on the rarefaction curves generated (Figure 6.1), were thought to account for most of the OTUs in the dataset. Sub-sampling resulted in the removal of all samples with sequence reads below 1000. The sub-sampled sequence reads were analysed in Mothur (Chapter 2, section 2.10.7) to improve their quality by removing PCR generated chimeric sequences and other sequencing errors, further decreasing the number of samples and OTUs present. Originally, the total number of OTUs identified was 173 OTUs contained in116 individuals. The combined process of sub-sampling and sequence cleaning resulted in the reduction of OTUs to 46 (73% reduction in number of OTUs), contained in 58 (43% reduction in sample size) subjects. Analysis of the 46 OTUs identified showed that 9 OTUs are associated with microorganisms frequently isolated from plant, soil or aquatic environments. A study in the Netherlands reported representatives of Cyanobacteria, which were thought to reflect plant chloroplasts obtained through inhalation, as well as environmental contaminants such as Deinococcus-Thermus, Nitrosopira, Planctomycetes and Chloroflexi (Bogaert et al., 2011). Other environmental contaminants have also frequently been detected as common reagent contaminants (Salter et al., 2014) which impact sequence-based microbiome analysis and were thus excluded from the original data: *Acidobacteria* (Ward et al., 2009); *Novosphingobium* (Sohn et al., 2004); *Caulobacter* (Hu et al., 2005); *Mesorhizobium* (Kaneko et al., 2000)*; Rhodospirillales*; *Bacillus licheniformis* (De Clerck and De Vos, 2004); *Bacillus niacini* (Nagel and Andreesen, 1991); *Acidovorax* (Shrestha et al., 2013, Schaad et al., 2008) ; *Dyella* (Weon et al., 2009); *Chitinophagacea; Sediminibacterium* (Kim et al., 2013); and most *Alphaproteobacteria* (Williams et al., 2007); *Xanthomonads* (Jones et al., 2000) and *Sphingobacteria*. Further analysis was performed on the

remaining 37 OTUs, which represented 21% (37/173) of all the OTUs detected. The total number of sequence reads per sample for the 37 OTUs was 1267, with OTU173 (70% (892/1267)) as the most dominant (Figure 6.3). As previously described, OTU173 was dominant and was detected in each sample in the entire dataset. The remaining 36 OTUs showed a very low representation at 30 % (375/1267) of the total number of reads for the 37 OTUs (Figure 6.3).

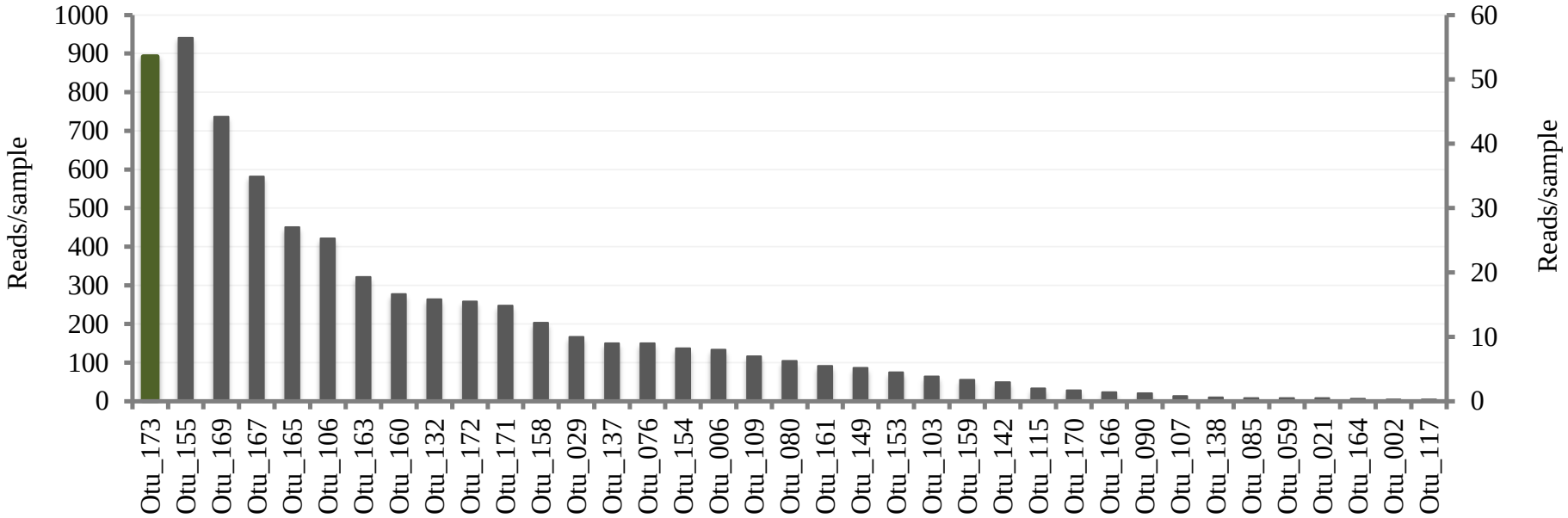


Figure 6.3 Sequence reads per sample in each OTU

The graph was generated using data from both adults and children. The graph represents 37 OTUs with an average of 1267.53 sequence reads per sample. The number of sequence reads per sample showed the abundance of each OTU in the dataset. The primary axis represents OTU173, which was the most abundant (70.4%; 892.36/1267.53). The remaining 36 OTUs are plotted using the secondary y-axis for clarity because of their low sequence reads per sample.

6.2.3. Classification of OTUs into taxonomic phyla

In this study, all the OTUs identified were classified into 4 taxonomic phyla namely: *Firmicutes* (79%), *Proteobacteria* (17%), *Actinobacteria* (3%), and *Bacteroidetes* (1%) (Figure 6.4). In brackets is the relative abundance of each phylum expressed as a percentage. These phyla have been reported in other settings (Bogaert et al., 2011, Lemon et al., 2010, Frank et al., 2010), however there is variation within the dominant phyla. For example, in the USA, the nasopharyngeal microbiota among healthy adults was dominated by *Actinobacteria* (68%), *Firmicutes* (27%), *Proteobacteria* (4%), *Bacteroidetes* (1.4%), and *Fusobacteria* (0.21%), however this shifted in hospitalised patients in the same setting with *Firmicutes* (71%) being the most abundant phyla followed by *Actinobacteria* (27%) (Frank et al., 2010). The increase in the abundance of *Firmicutes* in hospitalised patients was attributed to an increase in carriage of *S. aureus* and *S. epidermis* acquired from the hospital environment (Frank et al., 2010). This could also help explain the disproportionately high prevalence of *Firmicutes* in the Malawian setting because 86% (100/116) of all Malawian samples were collected from subjects who reported to Queen Elizabeth Central Hospital in Blantyre as outpatients. The remaining samples originated from infants in households in Karonga (Chapter 2, Section 2.1).

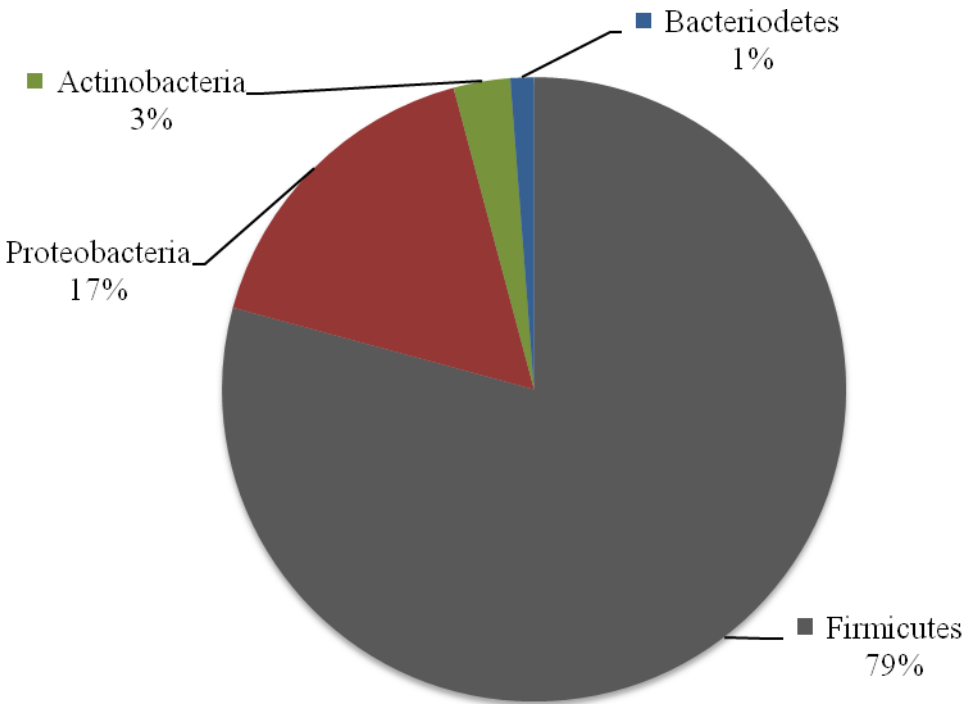


Figure 6.4 Taxonomic phyla and their relative abundance

The figure shows four phyla detected in the nasopharynx of Malawian children and adults. The most predominant phylum was *Firmicutes* followed by *Proteobacteria*, *Actinobacteria* and *Bacteroidetes.*

At the genus level (Table 6.2), the *Firmicutes* mainly consisted of the streptococcal species (73%), with *Staphylococcus*, *Bacillus*, *Gemella*, *Helcococus*, *Weissella* and *Lactobacillus* comprising a very small fraction (6%) of the total microbiota. *Proteobacteria*, which were the second most isolated phyla (17%) were dominated by *Acinetobacter* (4.45%) followed by *Moraxella* (3.18%), *Pseudomonas* (1.99%) *Klebsiella* (1.22%) and *Haemophilus* (0.89%). The remaining 5.27% comprised *Neisseria*, *Pantoea*, *Massilia*, *Curvibacter* and *Burkholderia*. Three genera were detected from the *Actinobacteria* phylum, the most predominant being

Corynebacterium (2.71%), followed by *Arthrobacter* (0.16%) and *Propionibacterium* (0.06%). *Bacteroidetes* were the least abundant phyla in the nasopharynx and these comprised two genera: *Porphyromonas* (1.18%) and *unclassified Flavobacteriacea* (0.04%). Compared to studies investigating nasopharyngeal microbiota in other settings (Bogaert et al., 2011, Chaban et al., 2013), there were similarities in the identities of the phyla and genera detected, however variations existed in the relative proportions of each phyla and dominant groups within each phyla.

Table 6.2 Profile of nasopharyngeal genera

A

Phyla	Class*	Order*	Family*	Genus	OTU	Reads/sample
Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus	OTU_173	70,40%
	Bacilli	Bacillales	Bacillaceae	Bacillus	OTU_169	3,49%
	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus	OTU_160	1,31%
	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus	OTU_132	1,25%
	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus	OTU_171	1,17%
	Bacilli	Lactobacillales	Camobacteriaceae	Dolosigranulum	OTU_137	0,71%
	Bacilli	Lactobacillales	Leuconostocaceae	Weissella	OTU_076	0,71%
	Bacilli	Bacillales	Staphylococcaceae	Gemella	OTU_090	0,10%
	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	OTU_059	0,04%
	Clostridia	Clostridiales	Peptostreptococcaceae	Helcococcus	OTU_164	0,03%
Actinobacteria	Actinobacteria	Actinomycetales	Corynebacteriaceae	Corynebacterium	OTU_165	2,13%
	Actinobacteria	Actinomycetales	Corynebacteriaceae	Corynebacterium	OTU_103	0,30%
	Actinobacteria	Actinomycetales	Corynebacteriaceae	Corynebacterium	OTU_159	0,26%
	Actinobacteria	Actinomycetales	Micrococcaceae	Arthrobacter	OTU_115	0,16%
	Actinobacteria	Actinomycetales	Propionibacterineae	Propionibacterium	OTU_107	0,06%
	Actinobacteria	Actinomycetales	Corynebacteriaceae	Corynebacterium	OTU_117	0,02%
Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Porphyromonas	OTU_006	0,63%
	Bacteroidia	Bacteroidales	Porphyromonadaceae	Porphyromonas	OTU_109	0,55%
	Flavobacteria	Flavobacteriales	Flavobacteriaceae	Unclassified	OTU_085	0,04%

B

Phyla	Class*	Order*	Family*	Genus	OTU	Reads/sample
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Acinetobacter	OTU_155	4,46%
	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Moraxella	OTU_167	2,75%
	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	OTU_106	1,99%
	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Acinetobacter	OTU_163	1,52%
	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Klebsiella	OTU_172	1,22%
	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	OTU_158	0,96%
	Gammaproteobacteria	Pasteurellales	Pasteurellaceae	Actinobacillus	OTU_029	0,79%
	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Pantoea	OTU_154	0,65%
	Betaproteobacteria	Neisseriales	Neisseriaceae	Neisseria	OTU_080	0,49%
	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Moraxella	OTU_161	0,43%
	Gammaproteobacteria	Pasteurellales	Pasteurellaceae	Haemophilus	OTU_149	0,41%
	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	Massilia	OTU_153	0,36%
	Gammaproteobacteria	Pasteurellales	Pasteurellaceae	Haemophilus	OTU_142	0,23%
	Gammaproteobacteria	Pasteurellales	Pasteurellaceae	Haemophilus	OTU_170	0,14%
	Gammaproteobacteria	Pasteurellales	Pasteurellaceae	Haemophilus	OTU166	0,11%
	Betaproteobacteria	Burkholderiales	Commonadaceae	Curvibacter	OTU_138	0,05%
	Betaproteobacteria	Neisseriales	Neisseriaceae	Neisseria	OTU_021	0,04%
	Betaproteobacteria	Burkholderiales	Burkholderiaceae	Burkholderia	OTU_002	0,02%

*The genera belonged to four different phyla: Firmicutes, Proteobacteria, Actinobacteria, and Bacteroidetes, presented in tables A and B. *The class, order and family for each genus were obtained from the NCBI taxonomy browser* (*http://www.nlm.nih.gov*).

6.3. Distribution of OTUs in children and adults

As previously described (Section 6.2.2), the general microbial diversity was analysed using samples that were subsampled to 1000 reads and with no duplicates for proper comparison of the whole dataset. This reduced the total number of samples from 116 to 58 samples. The 58 samples were from 58 subjects comprising children (n=40) and adults (n=18). The children were subdivided into HIV negative (n=21) and HIV positive (n=19). Adults were grouped into HIV negative (n=4) and HIV positive (n=14). From these samples, the total OTU richness was described by generating Venn diagrams in Mothur (Appendix 7). OTU richness defines the total number of OTUs present, while evenness measures the homogeneity of abundances in a given sample or environment (Bock et al., 2007, Thompson and Eithers, 2003). Results showed a total number (richness) of 38 OTUs in both adults and children (Figure 6.5). The total number of OTUs in children (n=40) was 28, which was less than the 34 OTUs detected in adults (n=18). The two cohorts however, had 23 OTUs in common (Figure 6.5). These findings suggest that age had an impact on the number of colonising microorganisms, with adults having more OTUs present than children. The diversity of OTUs in adults and children was further analysed using the inverse Simpson diversity index (Zhou et al., 2002). The diversity index measures microbial diversity by considering both richness (total number present) and evenness (level of diversity) and an increase in the inverse Simpsons diversity index corresponds to an increase in diversity (Zhou et al., 2002). Results showed no significant difference in the microbial diversity between children and adults ($P<0.4058$) (Figure 6.6), although this may not be definitive because of a limited sample size.

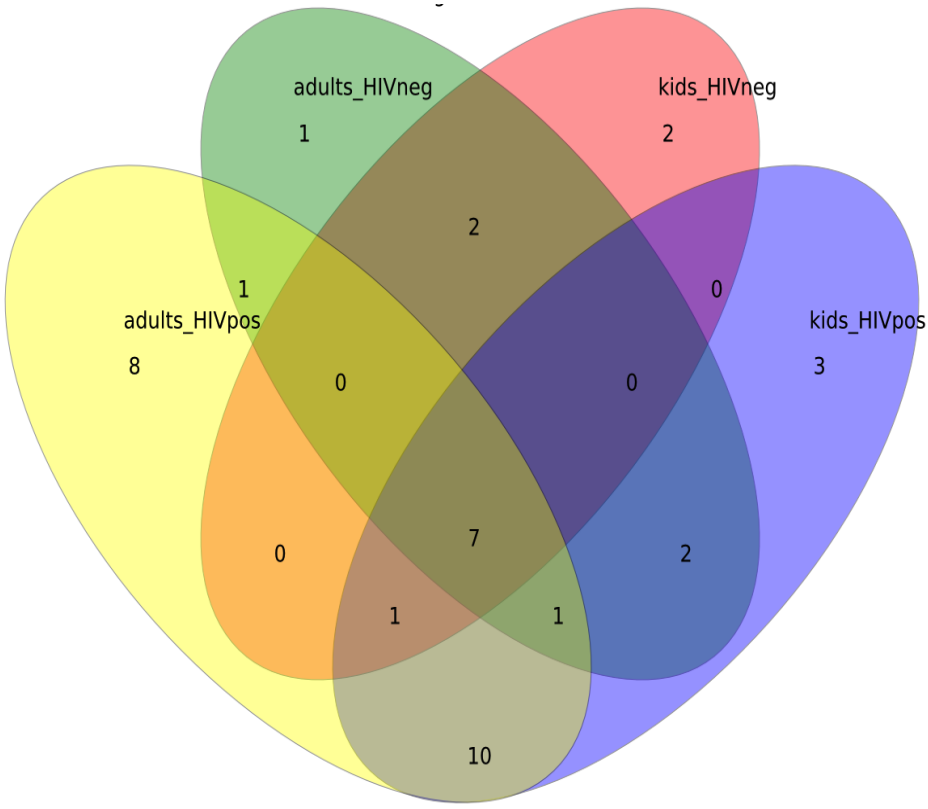


Figure 6.5 OTU richness in children and adults

The diagram was generated in Mothur to show the richness of microorganisms colonising adults (n=18) and children (n=40). In the figure, children are represented as ‘kids’.

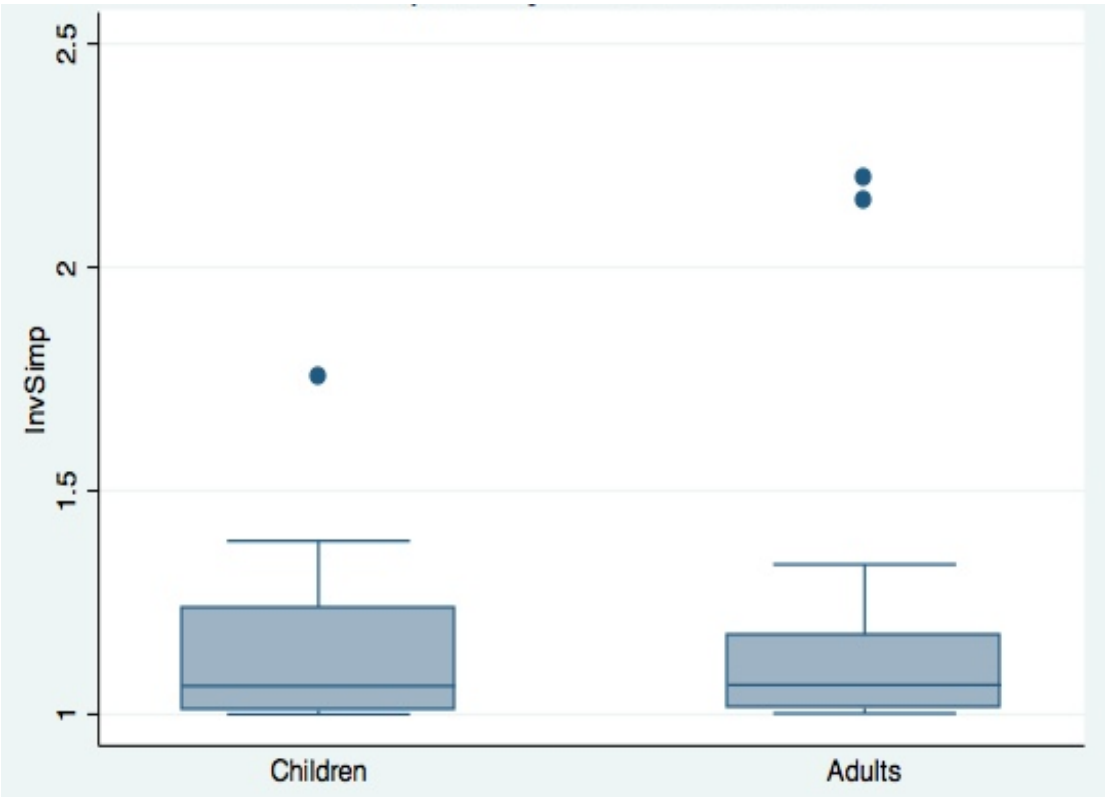


Figure 6.6 Inverse Simpson diversity index in children and adults

There was no significant difference in microbial diversity between children and adults ($P<0.4058$).

The microbial diversity in children and adults was further assessed using principal component analysis (PCA) (Figure 6.7). This was achieved using Bray-Curtis similarity cluster analysis or index of dissimilarity (Bray and Curtis, 1957) which describes the relative proportion of OTUs as well as their presence or absence in different communities. If two environments have the same microbial species composition; the Bray-Curtis index is assigned a value '0' and '1' when the microbial composition is completely different.

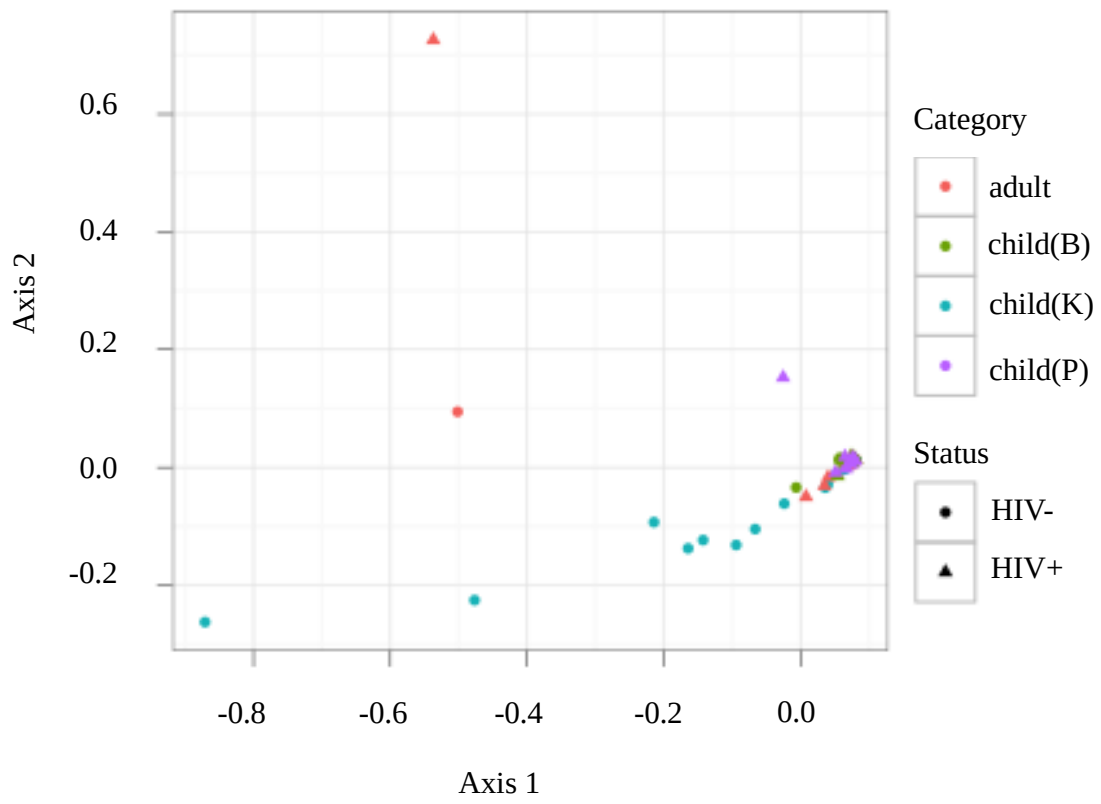


Figure 6.7 Bray-Curtis microbial diversity (PCA plots)

Principal component analysis plots presenting a two dimensional microbial diversity in adults and children by HIV status. The symbols correspond to individual subjects, clustered based on OTU diversity. This diversity is represented in a graph generated in mothur (Appendix 7) by statistically correlating every OTU with each axis (Axis 1 and Axis 2) on the PCA matrix. Children's samples were collected from Blantyre (Child (B) and Child (P)) and Karonga (Child (K)) while all adult samples were from Blantyre.

The results showed that the microbial composition in adults was very similar to each other (Figure 6.7) because the median values for the diversity score in both axes 1 and 2 were similar and close to the zero co-ordinates (Table 6.3). Children displayed a greater diversity in microbial composition amongst themselves, with samples

observed to scatter in the axis1 direction (Figure 6.7). This was also evident in the difference in the median diversity score for axis1: 0.25 (CI: -0.18-0.31) and axis2: -0.077 (CI: -0.19-0.08) (Table 6.3).

Table 6.3 Median diversity score in children and adults

	Median diversity (Interquartile range)	
Axis	Children	Adults
1	0.25 (-0.18-0.31)	-0.058 (-0.36-0.24)
2	-0.077 (-0.19-0.08)	0.087 (-0.14-0.26)

The table describes the median diversity score for children's and adults samples in both axes 1 and 2.

The scatter of the samples in the axis1 direction was caused by children's samples from Karonga, whilst samples from children in Blantyre belonged to the same cluster with adult samples. Using Spearman's rank correlation test in Mothur, the correlation coefficient for each OTU to the axes in the principal component analysis file was determined. Results showed that three OTUs: OTU173 (*Streptococcus*), OTU167 (*Moraxella*), and OTU165 (*Corynebacterium*) were largely responsible for the scatter in Axis1 direction; while OTU167 (*Moraxella*) and OTU160 (*Staphylococcus*) influenced the scatter in Axis2 in the PCA analysis. These findings suggest that there is high variation in microbial composition among children in Karonga, whereas microbial composition was less diverse and similar in children and adults from Blantyre.

6.4. Distribution of phyla in children and adults

Previously (Chapter 3, sections 3.2.1 and 3.4), a high diversity of pneumococcal serotype carriage was shown in Malawian adults and children, with no marked differences in the diversity between the two groups. In this chapter, we investigated whether the distribution of the four phyla detected was also similar in Malawian children and adults. The data was analysed in Mothur using Metastats (Appendix 8(a)), which is a statistical method designed to detect differentially abundant features between microbial communities i.e. features that are enriched or depleted in two different populations or environments (White et al., 2009).

A comparison of phyla diversity between children and adults. In our study, the four phyla: *Actinobacteria*, *Firmicutes*, *Proteobacteria* and *Bacteroidetes* were detected in both adults and children (Figure 6.8). Results showed no significant difference in the distribution of *Actinobacteria*, *Firmicutes* and *Proteobacteria* between children and adults (Table 6.4).

Table 6.4 Distribution of phyla in children and adults

Phyla	Child (mean)	Adult (mean)	p-value
Actinobacteria	0.022850	0.045389	0.790210
Bacteroidetes	0.000555	0.000834	0.028391
Firmicutes	0.899107	0.875984	0.732268
Proteobacteria	0.077488	0.077792	0.994006

However, *Bacteroidetes* were shown to be significantly higher in adults than children (P<0.028) (Table 6.4). *Bacteroidetes* comprised three OTUs (Table 6.2): OTU006 (*Porphyromonas*), OTU109 (*Porphyromonas*) and OTU085 (*Ornithobacterium*). There was no significant difference in the distribution of OTU006 (P<0.76) and OTU109 (P<0.48) between children and adults (Appendix 8 (a)) suggesting equal prevalence. However, OTU085 was significantly higher in adults than children (p<0.001), but further analysis showed that OTU085 was only isolated from a single adult sample (Figure 6.9), meaning the observed difference in the distribution of *Bacteroidetes* may not be a true reflection of differences at the population level. The impact of a single carrier observed in this analysis highlights the need for conducting such studies on a large dataset. Our findings however, are in agreement with a recent study, which demonstrated a shift in the microbiota of the human skin and nares with age (Oh et al., 2012), where a higher prevalence of *Firmicutes*, *Bacteroidetes*, and *Proteobacteria* was reported in the nares of children than adults.

A comparison of phyla diversity in children from Blantyre and Karonga. There was no significant difference in the distribution of *Bacteroidetes* in the two regions (p<0.52) (Table 6.5); however, the proportions of *Actinobacteria*, *Firmicutes* and *Proteobacteria* were significantly different (Table 6.5). The phyla *Actinobacteria* were significantly higher in children from Karonga than Blantyre (P<0.023) (Table 6.5). *Actinobacteria* comprised three genera: *Corynebacterium* (OTU165, OTU103, OTU159 and OTU117). The data demonstrated that there was a significantly high carriage prevalence of OTU 165 (p<0.0099) in children from Karonga, which may have contributed to the significantly higher carriage of *Actinobacteria* (P<0.023) in Karonga compared to Blantyre. All samples from Karonga were collected from HIV

negative children, and high carriage of *Corynebacterium* in Karonga is consistent with high carriage of *Corynebacterium* among HIV negative subjects.

Table 6.5 Distribution of phyla in children

Phyla	Karonga (mean)	Blantyre (mean)	p-value
Actinobateria	0.049751	0.004917	0.022977
Bacteroidetes	0.001137	0.000167	0.523477
Firmicutes	0.795831	0.967957	0.007992
Proteobacteria	0.153281	0.026959	0.015984

A comparison between children from Karonga and Blantyre.

In contrast, *Firmicutes* were significantly higher in children from Blantyre than those from Karonga (P<0.0079). By composition Firmicutes contained 10 OTUs (Table 6.2) but only three OTUs (OTU137, OTU160 and OTU173) showed a significant difference in the distribution between children in the two regions (Appendix 8(b)). OTU160 (*Staphylococcus aureus*) was significantly higher in children from Blantyre than Karonga (P<0.0099). Apart from differences in age and HIV status in the two groups, children's samples in Blantyre were collected from a hospital setting (Chapter 2, section 2.1) which might have increased their risk of exposure to carriage of *S. aureus* (Shibabaw et al., 2013), a commensal microorganism whose main ecological niche is the human nasopharynx (Kluytmans et al., 1997). Carriage of OTU173 (*Streptococcus*) was significantly higher in children from Blantyre than Karonga (P<0.024), suggesting high exposure to this genus in hospital among children in Blantyre. Thus, the high prevalence of *Firmicutes* in Blantyre is

influenced by the combined prevalence of *S. aureus* and *Streptococcus.* OTU137 (*Dolosigranulum pigrum*) was significantly higher in children from Karonga (P<0.00899) than Blantyre. *D.pigrum* was isolated from 6 out of 16 children in Karonga and none from Blantyre (Figure 6.9), suggesting localised distribution of this genus. *D. pigrum* is a Gram-positive catalase-negative coccus and a commensal inhabitant of the upper respiratory tract. It is reported to be a rare opportunistic pathogen causing non or ventilator associated pneumonia and septicaemia (Hoedemaekers et al., 2006, Lecuyer et al., 2007) and more recently corneal infection (Sampo et al., 2013).

The distribution of *Proteobacteria* was significantly higher in children from Karonga than Blantyre (P<0.015984). *Proteobacteria* was comprised of 18 OTUs; however only 3 OTUs namely OTU163 (*Acinetobacter*), OTU167 (*Moraxella*), and OTU172 (*Klebsiella)* showed a significant difference in their distribution between the two regions. OTU172 was only isolated from one child in Blantyre (Figure 6.14) so the observed significant difference is not be representative of the prevalence at the population level. OTU163 was isolated from 3 children in Blantyre and none in Karonga (Figure 6.14). *Acinetobacter* is common in hospital settings (Fournier and Richet, 2006, Munoz-Price and Weinstein, 2008) and this may explain why it was exclusively isolated form children in Blantyre whose samples were obtained during their visit to the hospital. *Acinetobacter* is a Gram-negative bacterium that is associated with contaminating environmental microorganisms (Tanner et al., 1998), however certain species are commensal of the human skin (Chu et al., 1999) and the nasopharynx (Anstey et al., 2002, Wolf et al., 2001). OTU167 represented *Moraxella*, which was isolated from children in both regions; however, it was highly

predominant and was detected in virtually all the children in Karonga ($P<0.002$). The children in Karonga were infants, much younger than the older children in Blantyre (≥2years). The age difference could play a role in the observed carriage prevalence of *Moraxella*, which has been reported to be higher in children (75%) than adults (1-3%) (Verduin et al., 2002). The genus *Moraxella* belongs to *Moraxellaceae* family comprising Gram-negative, oxidase and catalase positive coccobacilli; and commensally reside in the upper respiratory tract (Laukeland et al., 2002). These findings give an indication that there is a difference in the distribution of nasopharyngeal microbiota in children from the two regions, however, further analysis of samples of similar age and HIV status from the two regions would be recommended to establish whether indeed the observed differences could be attributed to differences in geographical location.

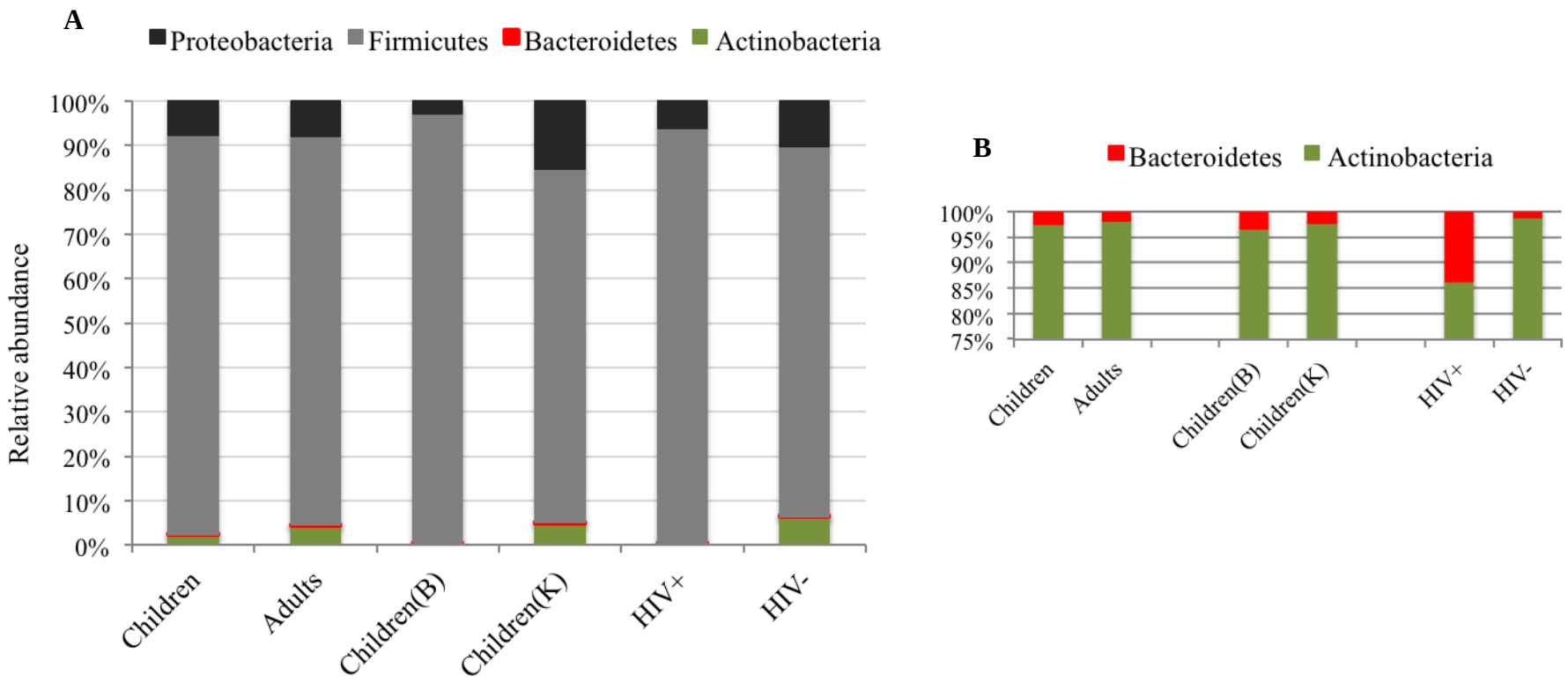


Figure 6.8 Distribution of Phyla in children and adults

Figure (A) shows the distribution of *Proteobacteria*, *Firmicutes*, *Bacteroidetes* and *Actinobacteria* in Malawian children and adults. Figure (B) shows distribution of *Bacteroidetes* and *Actinobacteria* for clarity because their relative abundance was very low.

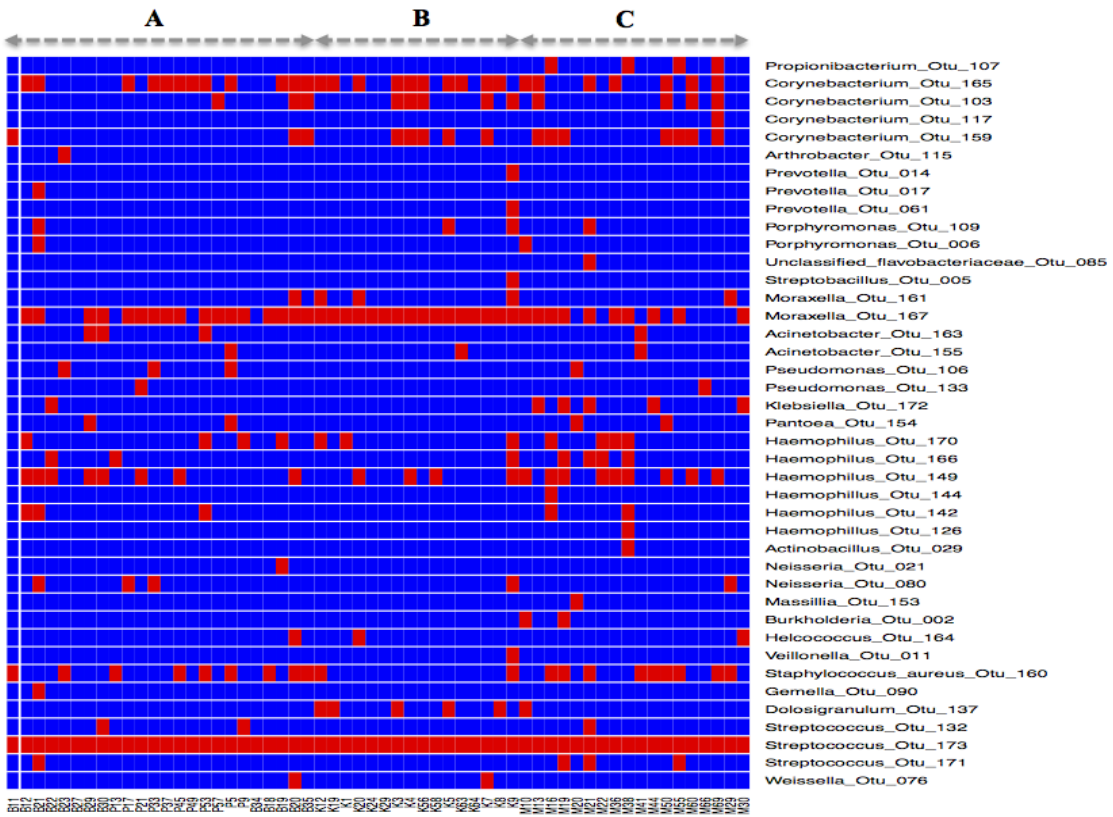


Figure 6.9 Distribution of microbial genera in children and adults

Sections ‘A’ and ‘B’ represent children from Blantyre and Karonga respectively, ‘C’ represents adults. The red squares signify carriage (presence) while the blue squares signify non-carriage (absence).

6.5. Distribution of OTUs and Phyla by HIV status

A high proportion of samples in this dataset were collected from HIV positive (n=35) compared to HIV negative (n=23) subjects. We investigated whether HIV infection had any impact on the distribution of OTUs and Phyla in carriage.

OTU distribution. The analysis identified 24 OTUs in HIV positive children (n=19), which was higher than 12 OTUs identified in HIV negative children (n=21) with 8 OTUs in common. However, the difference in OTU distribution was not statistically significant (P<0.0923). Again this may not be definitive because of the limited sample size. The total number of OTUs in HIV negative (n=4) and HIV positive (n=14) adults was 14 and 28 respectively, while 9 OTUs were shared (Figure 6.4). The difference in OTU distribution in adults was not statistically significant (P<0.3894). These results suggest that although OTU richness was higher in HIV infected adults and children, the differences observed were not statistically significant. Further analysis by the inverse Simpson diversity index however, showed that the microbial population was more diverse in HIV negative than HIV positive children and adults (P<0.0409) (Figure 6.10). This suggests that although HIV infected subjects had large number of OTU, they were mainly comprised of related species, leading to a decrease in microbial diversity following HIV infection. This is also supported by a previous study where a depletion of some commensal microorganisms in rectal mucosa among the HIV infected and replacement with a few opportunistic pathogens was observed (McHardy et al., 2013). It is not clear why HIV infected subjects had less microbial diversity compared to HIV negative individuals. The question is whether the compromised immune function in HIV positive individuals allows colonisation and spread of a particular group of microbial

species, which prevents others from attaching to the mucosal surface. Addressing this question would require further investigation.

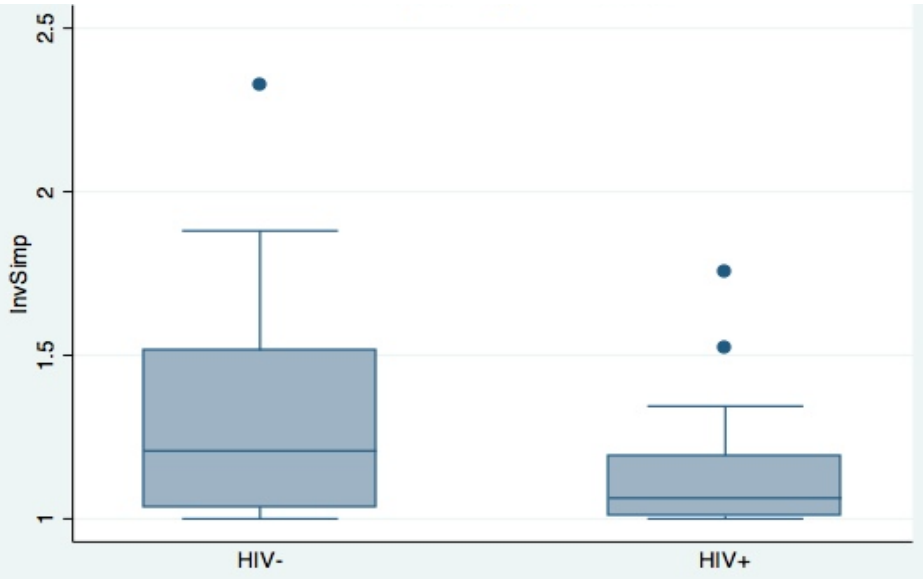


Figure 6.10 The inverse Simpson diversity index

There was higher microbial diversity in HIV negative than HIV positive adults and children (P<0.0409).

Phyla distribution. A shift from the normal profile of microbiota in various disease conditions has been documented (Costa et al., 2012, Cho and Blaser, 2012). HIV infection has been shown to be associated with changes in the microbial profiles amongst infected individuals (Saxena et al., 2012). In this study, the distribution of microbial species at the phylum and genus levels between HIV positive and HIV negative subjects was investigated. The aim was to establish if there would be a shift in nasopharyngeal carriage of different microbial species following HIV infection in Malawian children and adults. Results showed that all the four phyla (*Bacteroidetes,*

Firmicutes, Proteobacteria and Actinobacteria) were detected in both HIV positive and HIV negative subjects (Figure 6.8). There was no significant difference in carriage of *Bacteroidetes* (P<0.061), *Firmicutes* (P<0.088) and *Proteobacteria* (P<0.419) between the two cohorts, but the distribution of *Actinobacteria* was significantly higher in HIV negative than HIV positive subjects (P<0.007) (Table 6.6).

Table 6.6 Distribution of phyla by HIV status

Phyla	HIV +Ve (mean)	HIV -Ve (mean)	P-Value
Actinobacteria	0.003667	0.064400	0.006993
Bacteroidetes	0.000576	0.000728	0.061136
Firmicutes	0.935869	0.833932	0.088911
Proteobacteria	0.059887	0.100940	0.418581

At the genus level, *Actinobacteria* comprised *Corynebacterium* (OTU165, OTU103, OTU159 and OTU117); *Arthrobacater* (OTU115) and *Propionibacterium* (OTU107) (Table 6.2). There was no significant difference in the prevalence of *Arthrobacter* (P<0.621), *Propionibacterium* (P<0.424) and *Corynebacterium* (OTU159 (P<0.107), OTU117 (P<0.138)) between HIV positive and HIV negative subjects. However, OTU103 (*Corynebacterium accolens*) and OTU165 (*Corynebacterium pseudodiptheriticum*) were significantly reduced in HIV positive than HIV negative subjects, with p-values of 0.02 and 0.007 respectively (Appendix 8(c)). Thus OTU103 (*Corynebacterium accolens*) and OTU165 (*Corynebacterium pseudodipthericum*) were responsible for the significant shift in the prevalence of *Actinobacterium* by HIV status. The reason why the prevalence of *C.*

pseudodipthericum and *C. accolens* are reduced in HIV positive subjects is not clear, although competition with other co-colonising bacteria may influence diversity because *S. aureus* has been reported to associate negatively with *C. pseudodipthericum* and positively with *C. accolens*(Uehara et al., 2000). Our data however showed no significant difference in carriage of *S. aureus* (OTU 160) (Appendix 8 (c)) suggesting involvement of other host or environmental factors (Glennie et al., 2010).

6.6. Discussion

This study investigated the composition of nasopharyngeal microbiota of children (n=40) and adults (n=18) in the context of high prevalence of HIV infection and invasive pneumococcal disease. The aim was to determine if the distribution of the microbiota in the nasopharynx varied between adults and children and whether HIV infection resulted in a shift from the baseline microbiota in healthy individuals. The individuals were grouped based on their respective HIV serostatus; HIV negative (n=23) and HIV positive (n=35). The results showed 4 taxonomic phyla (*Actinobacteria*, *Bacteroidetes*, *Firmicutes*, *Proteobacteria*) and 37 species level phyla type (OTUs) were predominant in the Malawian samples. *Firmicutes* (79%) were the most abundant phyla, followed by *Proteobacteria* (17%), *Actinobacteria* (3%) and *Bacteroidetes* (1%). The types of phyla detected in Malawian adults and children are consistent with the findings in healthy children from the Netherlands (*Proteobacteria* (64%), *Firmicutes* (21%), *Bacteroidetes* (11%), *Actinobacteria* (3%), *Fusobacteria* (1.4%)) (Bogaert et al., 2011) and healthy adults from China (*Actinobacteria* (38.1%), *Firmicutes* (31.7%), *Proteobacteria* (5.4%), *Bacteroidetes* (1.4%)) (Ling et al., 2013). Other settings also reported similar composition: Egypt

in healthy adults (*Actinobacteria, Firmicutes and proteobacteria*) (Frank et al., 2010), Canada in pandemic H1N1 children and adults patients (*Actinobacteria, Firmicutes, Proteobacteria*) (Chaban et al., 2013) and USA in healthy adults (*Firmicutes, Actinobacteria*) (Lemon et al., 2010). It can be deduced from the above studies that *Proteobacteria, Firmicutes, Bacteroidetes and Fusobacteria* are the commonly isolated taxa from the nasopharynx of both children and adults across the world. These findings suggest that the human upper respiratory tract fundamentally shares a similar collection of commensal bacterial families across all age ranges. There are however variations in the dominant phyla or species based on the subjects state of health, age and geographical location (Garcia-Rodriguez and Fresnadillo Martinez, 2002). In Malawi, for example there was a disproportionately high representation of *Firmicutes* across all age ranges, compared to other settings. The reasons why Malawians had such a high abundance of *Firmicutes* are not clear. However, this could be influenced by a bias in the sampling of the Malawian dataset where only carriers of *S. pneumoniae* were analysed, as indicated previously (Chapter 2, Sections 2.1 and 2.2).

Diversity of nasopharyngeal microbiota by age. To investigate if age had an impact on nasopharyngeal microbiota, the profile of microbial species identified in children and adults was compared. Results showed that both children and adults carried the same profile of microbial species: *Firmicutes, Proteobacteria, Actinobacteria,* and *Bacteroidetes* which is consistent with other reports (Bogaert et al., 2011, Hilty et al., 2012). The profile of phyla between children and adults in Malawi was similar, however *Bacteroidetes* were significantly higher in adults than children ($P<0.028$). Elsewhere, prevalence of *Bacteroidetes* in the intestines has also

been reported to be significantly higher in adults than children (P<0.02) (Ringel-Kulka et al., 2013). This suggests an age dependent variation in the distribution of *Bacteroidetes*. The total richness was also found to be higher in adults than children (P<0.0085) although there was no difference in species diversity (P<0.4058), suggesting equal distribution of phyla in adults and children. This finding is similar to that of pneumococcal carriage, where no significant difference in diversity was observed between children and adults (Chapter 3). On the contrary a previous study in Canada, also conducted on a limited sample size (n=65) showed a trend towards increasing species diversity with age (Chaban et al., 2013). Their study was conducted on patients with the pandemic H1N1 influenza (Chaban et al., 2013), while our dataset had a high representation of samples from HIV infected subjects. The difference in phyla proportions between this and our study could also be attributed to the use of cpn60 (Chaban et al., 2013) and 16S rRNA genes as respective markers for bacterial identification. For studies using 16S rRNA, the choice of the variable region for comparing findings in different settings is important because no single variable region is adequate to identify microbial diversity in different environments (Chakravorty et al., 2007). The identities of nasopharyngeal phyla in studies using different 16S rRNA variable regions are similar however marked differences occur in their proportions (Chaban et al., 2013, Bogaert et al., 2011). Unlike the study in Canada, the estimation of diversity index in our dataset was difficult because it was largely influenced by carriage of one dominant OTU173 (streptococcal species) in both adults and children. This suggested that any differences in less abundant OTUs would not have a major impact on the overall population diversity.

The impact of HIV infection on the profile of nasopharyngeal microbiota. HIV infection has been reported to predispose patients to increased carriage and susceptibility to disease by mucosal bacteria (Pan et al., 2005). The increased carriage is attributed to the impairment of membrane integrity and function because of an influx of pro-inflammatory cytokines in the presence of the virus (Glennie et al., 2010). HIV infection and several other infections have widely been reported to cause a shift in the diversity of microbiota (McHardy et al., 2013, Costa et al., 2012, Clemente et al., 2012). In this study, we investigated whether HIV infection did indeed have an impact on the diversity of nasopharyngeal flora in Malawian children and adults. Results showed no significant difference in the total number of species carried by both HIV positive and HIV negative children and adults combined, however the species diversity was higher in HIV negative compared to HIV positive subjects ($P<0.0409$). Furthermore, the species *C. accolens* and *C. pseudodiptheriticum* from the phyla *Actinobacteria* were significantly reduced in HIV positive children and adults, with p-values of <0.02 and <0.007 respectively, consistent with previous findings on the impact of HIV infection on nasopharyngeal microbiota (McHardy et al., 2013). It is not quite clear why there is a reduction of these specific species in HIV infected subjects, however, the compromised immune function in the HIV infected individuals has been reported to alter carriage dynamics at the mucosal site (Glennie et al., 2010), which may increase the competition for colonisation. Antibiotic use has been reported to alter the profile of normal flora in carriage (Varon et al., 2000). Usage of co-trimoxazole prophylaxis in the HIV infected is recommended as it has been shown to reduce mortality (Nunn et al., 2008). Co-trimoxazole has also been reported to alter microbial carriage dynamics in the nasopharynx (Gill et al., 2008a, Arason et al., 1996). Therefore the use of co-

trimoxazole in Malawian HIV infected adults and children who were exclusively from Blantyre, may have contributed to the observed differences in microbial carriage dynamics in the two regions.

Microbial diversity in children. Environmental factors have been shown to play an important role in the diversity of the human microbiota (Levy and Borenstein, 2013). In this study, children's samples were collected from Blantyre and Karonga, over 400 miles apart (Chapter 2, section 2.0). Using a collectors curve, the species richness was found to be higher in children from Karonga than Blantyre (Figure 6.1). Analysis of species diversity by principal component analysis (Bray-curtis Dissimilarity Index) showed higher diversity in children from Karonga than Blantyre, suggesting a shift in the profile of nasopharyngeal microbiota in children from the two regions. The phyla detected in the two regions were the same and these included: *Actinobacteria, Bacteroidetes, Firmicutes and Proteobacteria*. Carriage of *Actinobacteria* was significantly higher in children from Karonga than Blantyre ($P<0.023$) and this was mainly due to the dominating *Corynebacterium species* ($P<0.0099$). Similarly, the carriage of *Proteobacteria* was significantly higher in children from Karonga than Blantyre ($P<0.015$), mainly due to *Moraxella* species ($P<0.002$). These results show *Corynebacterium* and *Moraxella* species were enriched in children from Karonga and were thus responsible for the increased prevalence of *Actinobacteria* and *Proteobacteria* in Karonga respectively. The distribution of *Dolosigranulum pigrum* was significantly higher in Karonga than Blantyre ($P<0.009$). It was further shown that *D. pigrum* was only localised to children in Karonga and none from Blantyre. *Firmicutes* were significantly higher in children from Blantyre than Karonga ($P<0.0079$), and this was mainly due to the

significantly high carriage of *Staphylococcus aureus* (P<0.0099) and *Streptococcus* (P<0.024). Overall *Staphylococcus aureus* and *Streptococcus* influenced the observed higher prevalence of *Firmicutes* in Blantyre than Karonga. These data show the shift in nasopharyngeal microbiota distribution between Karonga and Blantyre. All children in Karonga were infants (less than 12 months old) and exclusively HIV negative while the Blantyre samples were drawn from children under the age of five years and adults (including HIV positive individuals). However, a comparison in phyla distribution by HIV status in children was not made because of the dramatically reduced sample numbers after subsampling. The observed difference in microbial distribution in children could therefore arise from age differences in the two groups. Karonga samples were obtained from a rural setting whereas Blantyre samples were from the urban children reporting to hospital as outpatients. The difference in sampling source may also contribute to the diversity of microbial distribution in carriage.

6.7. Conclusion

In conclusion, findings from this study show that four phyla (*Firmicutes*, *Bacteroidetes, Actinobacteria* and *Bacteroidetes*) dominate nasopharyngeal microbiota in Malawian children and adults. No difference was observed in the distribution of these microorganisms in children and adults, suggesting no overall impact by age differences. However, there was a difference in microbial distribution between children from Karonga and Blantyre, which could possibly be attributed to HIV status because all samples from Karonga were HIV negative. *Actinobacteria* and *Proteobacteria* were dominant in Karonga whilst *Firmicutes* were dominant in Blantyre. Analysis of the impact of HIV infection on microbial distribution for the

whole dataset, showed a significant decrease in carriage of *Actinobacteria* ($P<0.007$). While the present study has characterised for the first time the nasopharyngeal microbiota in the Malawian population; and has also shown that HIV infection potentially influence carriage dynamics of the nasopharyngeal microbiota, there were limitations which made data interpretation difficult. The obvious limitation was a dramatic decrease in sample size following clean up and subsampling of the 454 sequences. The data also had a disproportionately high carriage of *Firmicutes*, which overshadowed the remaining three phyla making analysis of diversity scores difficult. Because of the important role microbiota plays in health and disease, further studies are recommended (Chapter 7, section 7.8) to address the current limitations.

Chapter 7: General discussion and conclusions

7.1 Introduction

This thesis has characterised nasopharyngeal carriage of *S. pneumoniae* and other microbiota in Malawian children and adults. Specifically, pneumococcal carriage and the prevalence of multiple pneumococcal serotype carriage were addressed (Chapter 1, Section 1.12). Carriage data is important in any setting for estimating impact of vaccination before and after it has been introduced (Goldblatt et al., 2013) and can be modelled to reflect changes occurring in invasive disease post vaccination (Weinberger et al., 2013). Carriage data have previously been generated for Malawi (Feikin et al., 2003) using Quellung and latex agglutination methods, however this was inadequate. The present study used microarray, which has the enhanced ability to determine carriage of multiple serotypes and detection of novel serotypes. By microarray, less abundant serotypes (usually missed by conventional Quellung and latex agglutination methods) were identified and this provided a good estimate of serotype diversity in Malawi. In addition, this data will form a good basis for estimating the impact of PCV13 introduction in Malawi. The last question addressed carriage of other nasopharyngeal microbiota in Malawian children and adults. These microbiota are involved in different modes of interaction which may impact on pneumococcal dynamics as well as virulence. The current study was not designed to characterise such interaction but identify for the first time to profile of nasopharyngeal microbiota in the Malawian setting.

7.2 Pneumococcal carriage in children and adults

Broad serotype diversity in Malawi children and adults. The study has demonstrated a broad range of pneumococcal serotypes and traces of nontypeable pneumococcal strains carried in Malawian children (Chapter 3, Section 3.2.1). Although pneumococcal carriage is generally higher in children than adults (Regev-Yochay et al., 2004b, Hussain et al., 2005, Mackenzie et al., 2010, Adetifa et al., 2012), our study showed that among carriers, pneumococcal serotype distribution was as equally diverse in adults (Chapter 3, Section 3.4) as it was in children. The high carriage diversity in adults could be attributed to the high representation of samples from HIV positive adults who were shown to carry a broader range of serotypes than HIV negative adults (Chapter 3, Section 3.4.1). HIV infection has been shown to significantly increase the risk of pneumococcal colonisation in adults (Gill et al., 2008b), which is thought to be associated with a defective pneumococcal specific Th1 response (Glennie et al., 2012a) and impairment of mucosal membrane barrier integrity (Glennie et al., 2010). PCV13 coverage of vaccine serotypes in Malawian children was estimated at 60% while in adults it was 30%, which is low compared to the 88% estimate in children and 50 % in adults based on strains from invasive disease in the same setting (Cornick et al., 2011, Everett et al., 2012). Such low estimate of PCV13 coverage highlights the higher diversity of serotypes in carriage, and suggests that carriage data may underestimates PCV13 coverage in a given population. Although carriage estimates for vaccine coverage were low, the fact that carriage is more common than disease (Brueggemann et al., 2003) it provides a good alternative for estimating vaccine efficacy. It is thought that the impact of any vaccine on carriage can be used as a guide for licensing and implementing new vaccines (Goldblatt et al., 2013), and recently it has also been

shown that modelling of carriage data can be used to estimate post-PCV changes in incidences of invasive pneumococcal disease in settings with no population-based surveillance systems for monitoring IPD incidences such as Malawi (Weinberger et al., 2013). This was achieved by estimating the change in carriage prevalence for each serotype in each study, followed by the calculation of the average change among vaccine and nonvaccine serotypes (Weinberger et al., 2013). These values were then multiplied by the proportion of IPD caused by each serotype during the pre-PCV period to obtain an estimate of post-PCV disease incidence (Weinberger et al., 2013). There were also a high proportion of non-vaccine serotypes (NVTs) circulating in children (40%) and adults (70%). It is well described (Weinberger et al., 2011, Mulholland and Satzke, 2012, Flasche et al., 2011b) that NVTs are responsible for serotype replacement post-vaccination and the eventual emergence of replacement disease. Serotype replacement in carriage has also been demonstrated in populations using PCV13 in the USA (Lee et al., 2013a) and Canada (Ricketson et al., 2014). Failure by PCV13 to reduce carriage of vaccine serotypes has also been demonstrated in children over two years of age (Lee et al., 2013a), and this can have a negative impact on herd immunity. The real impact of vaccination in a given population can only be assessed through a post-vaccination carriage and IPD surveillance, necessitating the need for maintaining on-going surveillance and the initiation of broader carriage studies post PCV13 introduction across Malawi.

High carriage of nontypeable serotypes in adults compared to children. Carriage of nontypeable *S. pneumoniae* accounted for 20% of all serotypes detected in adults and for 1.8% in children demonstrating a higher carriage of NT in adults than children. Higher carriage of NTs in adults is consistent with a study in Israel,

conducted on a much larger dataset, where NT represented 23.7% of all strains in adults and 5.2% in children (Regev-Yochay et al., 2004b). And a study from Thailand-Myanmar border area also showed a higher proportion of adults (mothers) carrying NT pneumococcal strains (29.6%) than infants (10.5%) using latex agglutination and Quellung reaction (Turner et al., 2012). These studies consistently show higher carriage of NT in adults compared to children. However, the reasons for this observation remain unclear. Reporting of NTs is often difficult because of their atypical colony morphology, which may misidentify NTs as non-pneumococcal streptococci (Turner et al., 2012). Therefore in reports using different serotyping methods, it is difficult to compare the findings and draw meaningful conclusions. In our study, it is extremely unlikely that NTs were misidentified, because the presence of pneumococcal strains was confirmed by microarray using PathID (Chapter 2, section 1.19). Our results therefore suggest that carriage of NTs in Malawi is more common among adults than children. A recent study has demonstrated high rates of genetic recombination occurring in nontypeable strains, and has also shown the occurrence of capsule switch event between encapsulated and nontypeable strains (Chewapreecha et al., 2014). Given the high rates of antibiotic resistance among NTs (Porat et al., 2006), the adult population will act as a source of genetic material that may provide a means to confer virulence and antibiotic resistance genes back to encapsulated pneumococcal strains.

High prevalence of multiple carriage in Malawian children compared to adults. This thesis has reported, for the first time, the prevalence of pneumococcal multiple pneumococcal serotypes carriage in Malawian children (Chapter 3, section 3.3) and adults (Chapter 3, section 3.5). Multiple carriage was detected on 40% (46/116)

across all the children's samples analysed and 19% (22/117) in adults (Chapter 3, section 3.5). This study was not able to establish the association between particular serotypes in multiple carriage because co-carriage of serotypes occurred at random. In adults, however, nontypeable serotypes were frequently isolated from multiple carriage with encapsulated pneumococcal strains more than any other serotype ($P<0.0058$) (Chapter 3, section 3.5). To the knowledge of the author, this is the first time the association between NTs and encapsulated pneumococcal strains has been identified in multiple carriage events. The association between NTs and encapsulated pneumococcal strains could not be established in children, possibly because of the extremely low prevalence of NTs in this cohort (1.8%). Multiple carriage has been reported in other settings, which allowed comparisons to be made with findings reported here. In The Gambia, multiple carriage was 19% (n=69) in children under the age of 2 years (Donkor et al., 2011), using microarray serotyping, which is low compared to Malawi despite using the identical molecular method. This difference can be attributed to differences in sampling. While the Gambian study estimated multiple carriage from analysing multiple colonies form the culture plate, The Malawian study used a plate sweep of all colonies (Chapter 2, section 2.5), which increased the chance of getting multiple serotypes. Similar to Malawian children, a study in Spain detected multiple carriage in 39% (41/105) of healthy children under the age of five years, by analysing a plate sweep of all colonies using multiplex PCR (Ercibengoa et al., 2012). In Thailand, a carriage study in children less than 2 years old (n=955) reported a low 12.9% prevalence of multiple carriage (Turner et al., 2013), by latex sweep serotyping, although this detected three times higher rates of multiple carriage compared to the standard WHO method (O'Brien and Nohynek, 2003) in the same study, suggesting that analysing all colonies from a plate culture

enhances multiple carriage detection. These studies reinforce the importance of methodology for accurate determination of carriage and multiple carriage in a given setting (Rapola et al., 1997). Other factors that have an impact on multiple carriage estimates include vaccination, which has been shown to lower multiple carriage in the vaccinated groups (Valente et al., 2012). This is unlikely in our dataset because the samples were collected prior to vaccine introduction in Malawi. However it would be interesting to monitor whether PCV13 would also reduce multiple carriage in a Malawian setting.

7.3 Impact of HIV infection on pneumococcal carriage

This thesis has shown that HIV negative children carry a broader range of serotypes compared to HIV positive children (Chapter 3, section 3.2.2). This suggests that HIV infection may result in a decrease in carriage serotype diversity in Malawian children. A recent study showed that serotype diversity in carriage was sensitive to the form of acquisition on nonspecific immunity and the strength of anti-capsular immunity. Since HIV infection compromises the hosts' immune system and function, it may explain why serotype diversity was higher in HIV negative than HIV positive children. In contrast, HIV positive adults carried a wider range of serotypes than HIV negative adults (Chapter 3, section 3.4.1) (Glennie et al., 2012a), however this was not statistically significant, suggesting the impact of HIV infection on carriage could not be demonstrated in adults. These results support recent findings in South Africa where carriage prevalence of pneumococcal serotypes was not different between HIV positive and HIV negative mothers (Nunes et al., 2013). Similarly another study also found no significant difference in the prevalence of pneumococcal serotypes between HIV positive and HIV negative adults (Rodriguez-Barradas et al., 1997).

The impact of HIV infection on carriage however is not fully understood because different studies have reported different findings, which highlights the need for further studies. In Kenya for example, there was no difference in the distribution of serotypes between the HIV-infected and population-based samples (Abdullahi et al., 2012a) in children under 2 years of age. In Zambia, HIV infection was reported as a risk factor for pneumococcal colonisation in mothers (Gill et al., 2008b). The impact of HIV infection on multiple carriage could not be demonstrated in both children and adults, which supports previous findings in Malawian adults (Glennie et al., 2012a). Different findings from different settings above, highlight the complexity of carriage in the context of HIV infection, as previously stated, and underscores the need for continued studies to fully understand how HIV infection influences carriage.

7.4 Carriage of capsular polysaccharide variants in Malawi

We have shown that genetic variations exist within the *CPS* loci of carried pneumococcal strains in both Malawian adults and children (Chapter 4). The predominant *CPS* variants belonged to serotype 6B, although variants of serotypes 19A, 20, and 29 were also detected. These variations occurred naturally i.e. without any vaccine selection pressure since samples were collected prior to the introduction of PCV13. Natural transformation of pneumococcal strains leading to capsule switching has previously been reported in Malawi (Everett et al., 2012). Variants of serotype 6B have previously been reported in the Netherlands (Elberse et al., 2011) and our variants were phylogenetically similar to 6B subtype III, although they exhibited different genotypes (Chapter 5, section 5.2). Such changes occurring within the *CPS* locus have the potential for compromising vaccine efficacy; as such it is not clear how the *CPS* locus variants observed in Malawi will respond towards

the vaccine. Further testing in mouse models to establish the ability of the variants to colonise, cause invasive disease or indeed escape the vaccine is currently underway. Preliminary findings suggest that the 6B variants colonise better than the wild type and are less invasive (unpublished), which is not surprising considering they were carriage serotypes. However, there are reports suggesting that these 6B variants could belong to the newly discovered serotype 6E (Ko et al., 2013). Further analysis of whole genomes and *CPS* gene variations is therefore required, to determine whether indeed the Malawian 6B variants belong to the recently discovered serotype 6E.

7.5 Nasopharyngeal microbiota in Malawian children and adults

Characterisation of the nasopharyngeal microbiota in Malawian adults and children showed four taxonomic phyla: *Firmicutes* (79%); *Proteobacteria* (17%); *Actinobacteria* (3%) and *Bacteroidetes* (1%) (Chapter6). Malawian adults and children demonstrated a disproportionately higher prevalence of *Firmicutes,* particularly streptococcal species in carriage than reported in other settings (Bogaert et al., 2011, Frank et al., 2010, Ling et al., 2013). Although the reasons for these observations are not clear, Malawian samples only included carriers of *S. pneumoniae*, which may have contributed to the disproportionately high prevalence of *Firmicutes*. However, to confirm whether indeed *Firmicutes* are the most predominant phyla in Malawi, further microbiome studies in both carriers and non-carriers of *S. pneumoniae* are recommended. A comparison of microbial diversity showed no variability in the distribution of phyla between adults and children in Malawi. Elsewhere, in Canada, a trend towards higher nasopharyngeal microbial diversity with age has been reported (Chaban et al., 2013). Unlike Malawi, where

16S rRNA was used as a target in a cohort with high HIV positive sample representation, the Canadian study (Chaban et al., 2013) used *cpn60* gene, as a universal target for microbial identification among the cohort diagnosed with H1N1 influenzae, which may contribute to the observed differences in microbial diversity. Other factors that have been reported to influence the apparent microbial diversity in different settings include contaminating DNA present in DNA extraction kits. Such contaminants are thought to critically impact the results, particularly those obtained from samples containing a low microbial biomass (Salter et al., 2014). Knowledge of common kit contaminants can enable comparisons of microbial diversity in settings using different reagents to be made. In this thesis, DNA contamination from kits was minimised by performing a blank DNA extraction containing all the reagents except the sample for each analysis. The impact of HIV infection on microbial diversity in Malawi was also investigated. Previous studies have shown that HIV infection causes a shift from the normal microbiota in humans (Saxena et al., 2012, McHardy et al., 2013). Our findings showed a significant decrease in carriage of *Actinobacteria* following HIV infection (P=0.007), although carriage of *Bacteroidetes* (P=0.061), *Firmicutes* (p=0.088) and *Proteobacteria* (p=0.419) remained unchanged. This suggests HIV infection does impact microbial carriage dynamics in the nasopharynx, however the ART mediated immune reconstitution did not have any impact on the richness and diversity of the nasopharyngeal microbiota in this dataset.

7.6 Study limitations

7.6.1 Sampling

Sample handling. The samples used for this study were retrieved from the MLW freezer archive, which is maintained at-80 °C in STGG (Chapter 2, section 2.4). To determine the microorganism present, the samples were thawed to 4 °C and then cultured on media. Culturing from frozen samples has been associated with loss in viability of certain pneumococcal serotypes, which for culture dependent methods, may affect the spectrum of serotypes obtained (Abdullahi et al., 2007) as well as underestimate the prevalence of multiple carriage. Analysing samples in the shortest possible time after collection to minimise the effect of freezing would be optimal. Employing culture independent molecular methods, which will improve detection of non-viable microorganisms, is also recommended.

Sample selection

1. Our study only analysed samples from nasopharyngeal swabs that were confirmed to contain pneumococcal strains. This likely introduced a bias in sample selection, which may have an impact on the reported distribution of pneumococcal strains, and microbiota in the nasopharynx. The bias can be minimised by including samples that are culture negative for *S. pneumoniae*in the analysis.

2. In order to understand the carriage dynamics in households overtime, a subset of samples was collected from infants at longitudinally. Samples from parents and older siblings were not available at the time of the study, which consequently made it difficult to fully understand the actual pneumococcal transmission dynamics in this setting. We suggest inclusion of other household members in future carriage

surveillance studies.

3. The longitudinal samples were only collected from four sampling points at four weeks interval, which is not an adequate for understand carriage dynamics in relation to duration of carriage and relationships among serotypes present in multiple carriage. We recommend sampling over a longer period of time.

7.6.2 Microarray serotyping

This thesis utilised microarray to characterise serotypes in carriage (Chapter 2, section 2.6). Unlike the Quellung reaction or latex agglutination, microarray has the enhanced utility to detect not only multiple carriage of *S. pneumoniae* and their relative abundance but it can also detect the presence of antibiotic resistance genes, novel serotypes and other colonising microbial species in the nasopharynx (Newton et al., 2011). Microarray also offers the advantage of analysing large number of samples within a short period of time and can be optimised to detect non-viable organisms (Satzke et al., 2014). The disadvantage of using microarray is that the reagents and equipment required are very costly, and demands a high level of technical expertise for array data analysis (Satzke et al., 2014). But most importantly, the array is not able to distinguish closely related serotypes because of identical sequences at the *CPS* locus. This limitation can be overcome by including more stringent targets for identification of each serotype (Satzke et al., 2014). Despite these limitations, microarray is still more sensitive than conventional methods and has wider applications (Turner et al., 2011).

7.6.3 Detection of nasopharyngeal microbiota

Limitations with 16S rRNA. The 16S rRNA gene (1550bp long) is universal in all bacteria, and has been widely used to estimate phylogenetic relationships among bacterial species (Clarridge, 2004). In clinical settings, the 16S rRNA gene has been used to accurately identify organism misidentified by conventional phenotypic methods (Petti et al., 2005). Despite the increasing use in clinical and other settings, the16S rRNA approach has some limitations:

1. The 16S rRNA gene sequencing can only classify bacteria up to genus level and there is currently no universal definition for the identification of species through 16S rRNA gene sequencing (Janda and Abbott, 2007).

2. Failure to distinguish closely related species for example, the 16S rRNA genes for species such as *Bacillus globisporus* and *Bacillus psychropilus* share over 99.5% sequence similarity, however the whole genome DNA homology is less than 50% (Fox et al., 1992). This suggests that the sequence similarity by 16S rRNA does not necessarily imply accurate identification.

3. The accurate identification of bacteria by 16S rRNA sequences is as good as the quality and completeness of the nucleotide sequences deposited into the public database, however the database has been reported to be incomplete and has sequencing errors (Boivin-Jahns et al., 1995, Ashelford et al., 2005).

Limitations with 454 sequencing: To characterise nasopharyngeal microbiota in our cohort, our study employed the 454 pyrosequencing (Chapter 2, section 2.11.6)

by sequencing the PCR amplified 16S rRNA gene. Although 454 sequencing has longer read length than the others and operates as a high-throughput sequencing tool (Rothberg and Leamon, 2008), it is more expensive per base and requires a substantial amount of starting material compared to other platforms such Illumina, Genome Analyser and Applied Biosystems (Roh et al., 2010). The major limitation facing 454 pyrosequencing and other NGS platforms is the production of sequences that are artefacts of PCR (Qiu et al., 2001) and sequencing (Quince et al., 2011, Schloss et al., 2011). These inflate the number of operational taxonomic units in a given sample, giving rise to erroneous results. The other problem is failure to accurately call long homopolymer bases, and the accuracy of base calling has been shown to decrease with increase in the length of the homopolymer, which are wrongly reported as insertions or deletions (Quinlan et al., 2008). For accurate determination of microbial species, it is necessary to improve the quality of the sequences generated by identifying and removing the sequence artefacts. To resolve these problems, bioinformatics tools are employed (Quince et al., 2011, Schloss et al., 2011). This study employed Mothur, a single piece of software, built on several pre-existing software for describing and comparing microbial communities (Schloss et al., 2009), in Malawian adults and children.

7.7 Conclusions

The study, on which this thesis is based, characterised the nasopharyngeal carriage of *S. pneumoniae* and other microbiota in Malawian children and adults. The results demonstrated that Malawian children and adults possess a diverse distribution of pneumococcal serotypes in carriage. The PCV13 coverage estimates was higher in children (60%) than adults (30%), suggesting higher carriage of vaccine serotypes in

children and supports the rationale behind the introduction of introducing PCV13 to children in Malawi. In addition, the data have also shown that carriage of non-vaccine serotypes (NVTs) was also very high in Malawian children and adults, suggesting the NVTs may act as reservoir for serotype replacement post-vaccination. Serotype replacement becomes a major clinical issue if it leads to NVT disease, which may override the benefits of vaccine introduction. For the first time in Malawi, the data have shown that multiple carriage is not only common but is more prevalent in children (40%) than adults (19%). Multiple carriage provides an ideal environment for horizontal gene transfer that drives pneumococcal evolution events such as acquisition of virulence and antibiotic resistance genes as well as capsule switch. There is evidence of genetic changes within the CPS loci of both vaccine and non-vaccine serotypes occurring naturally in the Malawian population, possibly due to high multiple carriage prevalence. However, this may accelerate under vaccine selection pressure and may have a negative impact on treatment. Malawi is a setting with high prevalence of HIV infection and invasive disease. While the prevalence of pneumococcal carriage is also high, the impact of HIV on carriage has never been investigated in Malawi. Generally, the data presented in this thesis has demonstrated no significant impact by HIV infection on the distribution of single or multiple pneumococcal serotypes carriage in both children and adults. Finally, the study investigated carriage of other microbiota colonising the nasopharynx because their presence could heavily impact on pneumococcal carriage dynamics as well as its virulence. The data have shown that *Firmicutes, Proteobacteria, Actinobacteria* and *Bacteroidetes* dominate carriage in the nasopharynx of Malawian adults and children, and there was no significant difference in the distribution of these phyla by age. However, carriage of *Actinobacteria* was significantly reduced in the HIV

infected. Although new data is presented here, providing new insights and scope for the future work, carriage still remains a complex web of interactions between host and the commensal microbial population. It requires further investigation to untangle its dynamics and build on the data generated in this thesis.

7.8 Future work

The carriage data presented in this thesis has provided valuable information on pneumococcal serotype distribution, and also provided preliminary data on the prevalent microbiota in carriage in Malawian children and adults. However the findings from this study do not only highlight the need to address many questions raised with a larger and better-matched sample sets, but also provide the basis for estimating much more accurately the size of the samples needed.

Surveillance of pneumococcal serotypes in carriage. In this study, sampling for determining carriage of pneumococcal serotypes in both adults and children was conducted before the introduction of pneumococcal conjugate vaccination in Malawi. Therefore the data presented in this study provides good pre-vaccine background information for estimating vaccine coverage and also act as a basis for determining how serotype profile changes in carriage post-vaccination. In Malawi, PCV13 was only introduced to children below the age of 2 years. To monitor the direct impact of vaccination in this group, further surveillance of pneumococcal strains in carriage amongst the vaccinated children would be recommended. Since PCV has also been shown to reduce carriage of vaccine serotypes among the unvaccinated persons through herd immunity, there should be expanded surveillance of pneumococcal carriage in the unvaccinated children and adults, to monitor this impact.

Dynamics of pneumococcal carriage over-time. To fully understand the dynamics of pneumococcal carriage in Malawian children overtime, further work is recommended. In particular, there would be need to sample mother/infant pairs over a longer period of time. This will provide the information on carriage dynamics as well as an estimate of carriage duration.

Investigating the ability of *CPS* locus variants to colonise and cause disease. The work presented in this thesis has shown that carriage serotypes in Malawi are genetically diverse. In particular, there is natural occurrence of genetic variation within the *CPS* loci of both vaccine and non-vaccine carriage serotypes. Such variations may generate serotypes, which are resistant to vaccination, and also modulate the virulence of other serotypes, which rarely cause invasive disease. To establish the real impact of such genetic changes on the phenotype of pneumococcal strains, further work is planned as follows:

1. Characterise whole genomes of *CPS* locus variants and comparing them to wild type serotypes in order to reveal differences other than those detected at the *CPS* locus; in particular focusing on genes that encode for virulent pneumococcal surface proteins.

2. Investigate the ability of the variants to colonise and cause invasive disease using mouse models. This work is currently in progress in our research group beginning with serotype 6B as a vaccine serotype as one of the most commonly isolated serotype from carriage and invasive disease in Malawi.

3. Investigating the response of the variants towards vaccination with conjugate vaccines. This will initially be achieved using mouse models as above, and secondly, through a post-PCV surveillance carriage study to determine any changes in the prevalence of the variants, since conjugate vaccination has been reported to decrease the carriage of vaccine serotypes.

Nasopharyngeal microbiome. To clearly characterise microbial diversity in carriage in Malawi, further analysis of nasopharyngeal microbiota is recommended. The samples used to characterise the nasopharyngeal microbiota were preselected based on carriage of *S. pneumoniae*. The results reported in this study showed extremely high carriage of streptococcal species belonging to *Firmicutes,* which made the estimate of diversity scores difficult. Further work on nasopharyngeal microbiota in this setting should include samples from both carriers and non-carriers of *S. pneumoniae,* to avoid over representation by one species.

Publications

1. **Kamng'ona, A.W**., Hinds, J., Bar-Zeev, N., Gould, K., Chaguza, C., Msefula, C., Cornick, J.E., Kulohoma, B.W., Gray, K., Bentley, S.D., French, N., Heyderman, R.S., Everett, D.B. 2014.High multiple-carriage of diverse pneumococcal serotypes and genetic variation at the capsular polysaccharide locus of vaccine serotypes in Malawian children. *Manuscript submitted to the Journal of Clinical Microbiology*. [Submitted for publication]

2. Cornick, J. E., Harris, S. R., Parry, C. M., Moore, M. J., Jassi, C., **Kamng'ona, A**., Kulohoma, B., Heyderman, R. S., Bentley, S. D. & Everett, D. B. 2013. Genomic identification of a novel co-trimoxazole resistance genotype and its prevalence amongst Streptococcus pneumoniae in Malawi. *The Journal of Antimicrobial Chemotherapy*, 69, 2, 368-74.

3. Kulohoma, B. W., Gray, K., **Kamng'ona, A**., Cornick, J., Bentley, S. D., Heyderman, R. S., Everett, D. B. 2013. Piliation of invasive *Streptococcus pneumoniae* isolates in the era prior to pneumococcal conjugate vaccine introduction, in Malawi. *Clinical and Vaccine Immunology*, 20, 11, 1729-35.

4. Glennie, S. J., Banda, D., Gould, K., Hinds, J., **Kamng'ona, A**., Everett, D. D., Williams, N. A. & Heyderman, R. S. 2012a. Defective Pneumococcal-Specific Th1 Responses in HIV-Infected Adults Precedes a Loss of Control of Pneumococcal Colonization. *Clinical Infectious Diseases*, 56, 2, 291-9.

References

2007. Pneumococcal Conjugate Vaccine For Childhood Immunization. WHO Position Paper. *Wkly Epidemiol Rec,* 82**,** 93-104.

2008. 23-Valent Pneumococcal Polysaccharide Vaccine. WHO Position Paper. *Wkly Epidemiol Rec,* 83**,** 373-84.

Aanensen, D. M., Mavroidi, A., Bentley, S. D., Reeves, P. R. & Spratt, B. G. 2007. Predicted Functions And Linkage Specificities Of The Products Of The Streptococcus Pneumoniae Capsular Biosynthetic Loci. *J Bacteriol,* 189**,** 7856-76.

Aanensen, D. M. & Spratt, B. G. 2005. The Multilocus Sequence Typing Network: Mlst.Net. *Nucleic Acids Res,* 33**,** 728-33.

Abdullahi, O., Karani, A., Tigoi, C. C., Mugo, D., Kungu, S., Wanjiru, E., Jomo, J., Musyimi, R., Lipsitch, M. & Scott, J. A. 2012a. The Prevalence And Risk Factors For Pneumococcal Colonization Of The Nasopharynx Among Children In Kilifi District, Kenya. *Plos One,* 7**,** E30787.

Abdullahi, O., Karani, A., Tigoi, C. C., Mugo, D., Kungu, S., Wanjiru, E., Jomo, J., Musyimi, R., Lipsitch, M. & Scott, J. A. 2012b. Rates Of Acquisition And Clearance Of Pneumococcal Serotypes In The Nasopharynges Of Children In Kilifi District, Kenya. *The Journal Of Infectious Diseases,* 206**,** 1020-9.

Abdullahi, O., Nyiro, J., Lewa, P., Slack, M. & Scott, J. A. 2008a. The Descriptive Epidemiology Of Streptococcus Pneumoniae And Haemophilus Influenzae

Nasopharyngeal Carriage In Children And Adults In Kilifi District, Kenya. *The Pediatric Infectious Disease Journal,* 27**,** 59-64.

Abdullahi, O., Nyiro, J., Lewa, P., Slack, M. & Scott, J. A. 2008b. The Descriptive Epidemiology Of Streptococcus Pneumoniae And Haemophilus Influenzae Nasopharyngeal Carriage In Children And Adults In Kilifi District, Kenya. *Pediatr Infect Dis J,* 27**,** 59-64.

Abdullahi, O., Wanjiru, E., Musyimi, R., Glass, N. & Scott, J. A. 2007. Validation Of Nasopharyngeal Sampling And Culture Techniques For Detection Of Streptococcus Pneumoniae In Children In Kenya. *J Clin Microbiol,* 45**,** 3408-10.

Ada, G. 2001. Vaccines And Vaccination. *N Engl J Med,* 345**,** 1042-53.

Adamou, J. E., Heinrichs, J. H., Erwin, A. L., Walsh, W., Gayle, T., Dormitzer, M., Dagan, R., Brewah, Y. A., Barren, P., Lathigra, R., Langermann, S., Koenig, S. & Johnson, S. 2001. Identification And Characterization Of A Novel Family Of Pneumococcal Proteins That Are Protective Against Sepsis. *Infect Immun,* 69**,** 949-58.

Adegbola, R. A., Hill, P. C., Secka, O., Ikumapayi, U. N., Lahai, G., Greenwood, B. M. & Corrah, T. 2006. Serotype And Antimicrobial Susceptibility Patterns Of Isolates Of Streptococcus Pneumoniae Causing Invasive Disease In The Gambia 1996-2003. *Trop Med Int Health,* 11**,** 1128-35.

Adetifa, I. M., Antonio, M., Okoromah, C. A., Ebruke, C., Inem, V., Nsekpong, D., Bojang, A. & Adegbola, R. A. 2012. Pre-Vaccination Nasopharyngeal

Pneumococcal Carriage In A Nigerian Population: Epidemiology And Population Biology. *Plos One,* 7**,** E30548.

Ahmed, S., Macfarlane, G. T., Fite, A., Mcbain, A. J., Gilbert, P. & Macfarlane, S. 2007. Mucosa-Associated Bacterial Diversity In Relation To Human Terminal Ileum And Colonic Biopsy Samples. *Appl Environ Microbiol,* 73**,** 7435-42.

Albrich, W. C., Madhi, S. A., Adrian, P. V., Van Niekerk, N., Mareletsi, T., Cutland, C., Wong, M., Khoosal, M., Karstaedt, A., Zhao, P., Deatly, A., Sidhu, M., Jansen, K. U. & Klugman, K. P. 2012. Use Of A Rapid Test Of Pneumococcal Colonization Density To Diagnose Pneumococcal Pneumonia. *Clinical Infectious Diseases : An Official Publication Of The Infectious Diseases Society Of America,* 54**,** 601-9.

Almeida, S. T., De Lencastre, H. & Sa-Leao, R. 2013. Epidemiology And Population Structure Of Serotypes 1, 5 And 7f Carried By Children In Portugal From 1996-2010 Before Introduction Of The 10-Valent And 13-Valent Pneumococcal Conjugate Vaccines. *Plos One,* 8**,** E75442.

Almirall, J., Bolibar, I., Serra-Prat, M., Roig, J., Hospital, I., Carandell, E., Agusti, M., Ayuso, P., Estela, A. & Torres, A. 2008. New Evidence Of Risk Factors For Community-Acquired Pneumonia: A Population-Based Study. *Eur Respir J,* 31**,** 1274-84.

Aniansson, G., Alm, B., Andersson, B., Larsson, P., Nylen, O., Peterson, H., Rigner, P., Svanborg, M. & Svanborg, C. 1992. Nasopharyngeal Colonization During

The First Year Of Life. *The Journal Of Infectious Diseases,* 165 Suppl 1**,** S38-42.

Anstey, N. M., Currie, B. J., Hassell, M., Palmer, D., Dwyer, B. & Seifert, H. 2002. Community-Acquired Bacteremic Acinetobacter Pneumonia In Tropical Australia Is Caused By Diverse Strains Of Acinetobacter Baumannii, With Carriage In The Throat In At-Risk Groups. *J Clin Microbiol,* 40**,** 685-6.

Arason, V. A., Kristinsson, K. G., Sigurdsson, J. A., Stefansdottir, G., Molstad, S. & Gudmundsson, S. 1996. Do Antimicrobials Increase The Carriage Rate Of Penicillin Resistant Pneumococci In Children? Cross Sectional Prevalence Study. *Bmj,* 313**,** 387-91.

Arbique, J. C., Poyart, C., Trieu-Cuot, P., Quesne, G., Carvalho Mda, G., Steigerwalt, A. G., Morey, R. E., Jackson, D., Davidson, R. J. & Facklam, R. R. 2004a. Accuracy Of Phenotypic And Genotypic Testing For Identification Of Streptococcus Pneumoniae And Description Of Streptococcus Pseudopneumoniae Sp. Nov. *Journal Of Clinical Microbiology,* 42**,** 4686-96.

Arbique, J. C., Poyart, C., Trieu-Cuot, P., Quesne, G., Carvalho Mda, G., Steigerwalt, A. G., Morey, R. E., Jackson, D., Davidson, R. J. & Facklam, R. R. 2004b. Accuracy Of Phenotypic And Genotypic Testing For Identification Of Streptococcus Pneumoniae And Description Of Streptococcus Pseudopneumoniae Sp. Nov. *J Clin Microbiol,* 42**,** 4686-96.

Ardanuy, C., Tubau, F., Linares, J., Dominguez, M. A., Pallares, R. & Martin, R. 2005. Distribution Of Subclasses Mefa And Mefe Of The Mefa Gene Among Clinical Isolates Of Macrolide-Resistant (M-Phenotype) Streptococcus

Pneumoniae, Viridans Group Streptococci, And Streptococcus Pyogenes. *Antimicrobial Agents And Chemotherapy,* 49**,** 827-9.

Arnold, R. R., Brewer, M. & Gauthier, J. J. 1980. Bactericidal Activity Of Human Lactoferrin: Sensitivity Of A Variety Of Microorganisms. *Infection And Immunity,* 28**,** 893-8.

Arpin, C., Daube, H., Tessier, F. & Quentin, C. 1999. Presence Of Mefa And Mefe Genes In Streptococcus Agalactiae. *Antimicrobial Agents And Chemotherapy,* 43**,** 944-6.

Arrecubieta, C., Lopez, R. & Garcia, E. 1994. Molecular Characterization Of Cap3a, A Gene From The Operon Required For The Synthesis Of The Capsule Of Streptococcus Pneumoniae Type 3: Sequencing Of Mutations Responsible For The Unencapsulated Phenotype And Localization Of The Capsular Cluster On The Pneumococcal Chromosome. *J Bacteriol,* 176**,** 6375-83.

Asahi, Y., Yoshikawa, T., Watanabe, I., Iwasaki, T., Hasegawa, H., Sato, Y., Shimada, S., Nanno, M., Matsuoka, Y., Ohwaki, M., Iwakura, Y., Suzuki, Y., Aizawa, C., Sata, T., Kurata, T. & Tamura, S. 2002. Protection Against Influenza Virus Infection In Polymeric Ig Receptor Knockout Mice Immunized Intranasally With Adjuvant-Combined Vaccines. *J Immunol,* 168**,** 2930-8.

Ashelford, K. E., Chuzhanova, N. A., Fry, J. C., Jones, A. J. & Weightman, A. J. 2005. At Least 1 In 20 16s Rrna Sequence Records Currently Held In Public Repositories Is Estimated To Contain Substantial Anomalies. *Appl Environ Microbiol,* 71**,** 7724-36.

Attali, C., Durmort, C., Vernet, T. & Di Guilmi, A. M. 2008a. The Interaction Of Streptococcus Pneumoniae With Plasmin Mediates Transmigration Across Endothelial And Epithelial Monolayers By Intercellular Junction Cleavage. *Infection And Immunity,* 76**,** 5350-6.

Attali, C., Frolet, C., Durmort, C., Offant, J., Vernet, T. & Di Guilmi, A. M. 2008b. Streptococcus Pneumoniae Choline-Binding Protein E Interaction With Plasminogen/Plasmin Stimulates Migration Across The Extracellular Matrix. *Infection And Immunity,* 76**,** 466-76.

Austrian, R. & Collins, P. 1966. Importance Of Carbon Dioxide In The Isolation Of Pneumococci. *Journal Of Bacteriology,* 92**,** 1281-4.

Balachandran, P., Brooks-Walter, A., Virolainen-Julkunen, A., Hollingshead, S. K. & Briles, D. E. 2002. Role Of Pneumococcal Surface Protein C In Nasopharyngeal Carriage And Pneumonia And Its Ability To Elicit Protection Against Carriage Of Streptococcus Pneumoniae. *Infect Immun,* 70**,** 2526-34.

Basset, A., Thompson, C. M., Hollingshead, S. K., Briles, D. E., Ades, E. W., Lipsitch, M. & Malley, R. 2007. Antibody-Independent, Cd4+ T-Cell-Dependent Protection Against Pneumococcal Colonization Elicited By Intranasal Immunization With Purified Pneumococcal Proteins. *Infect Immun,* 75**,** 5460-4.

Bates, S. T., Berg-Lyons, D., Caporaso, J. G., Walters, W. A., Knight, R. & Fierer, N. 2011. Examining The Global Distribution Of Dominant Archaeal Populations In Soil. *The Isme Journal,* 5**,** 908-17.

Bentley, S. D., Aanensen, D. M., Mavroidi, A., Saunders, D., Rabbinowitsch, E., Collins, M., Donohoe, K., Harris, D., Murphy, L., Quail, M. A., Samuel, G., Skovsted, I. C., Kaltoft, M. S., Barrell, B., Reeves, P. R., Parkhill, J. & Spratt, B. G. 2006. Genetic Analysis Of The Capsular Biosynthetic Locus From All 90 Pneumococcal Serotypes. *Plos Genet,* 2**,** E31.

Bera, A., Biswas, R., Herbert, S. & Gotz, F. 2006. The Presence Of Peptidoglycan O-Acetyltransferase In Various Staphylococcal Species Correlates With Lysozyme Resistance And Pathogenicity. *Infection And Immunity,* 74**,** 4598-604.

Berry, A. M. & Paton, J. C. 1996. Sequence Heterogeneity Of Psaa, A 37-Kilodalton Putative Adhesin Essential For Virulence Of Streptococcus Pneumoniae. *Infect Immun,* 64**,** 5255-62.

Black, S., Shinefield, H., Fireman, B., Lewis, E., Ray, P., Hansen, J. R., Elvin, L., Ensor, K. M., Hackell, J., Siber, G., Malinoski, F., Madore, D., Chang, I., Kohberger, R., Watson, W., Austrian, R. & Edwards, K. 2000. Efficacy, Safety And Immunogenicity Of Heptavalent Pneumococcal Conjugate Vaccine In Children. Northern California Kaiser Permanente Vaccine Study Center Group. *Pediatr Infect Dis J,* 19**,** 187-95.

Bock, C. E., Jones, Z. F. & Bock, J. H. 2007. Relationships Between Species Richness, Evenness, And Abundance In A Southwestern Savanna. *Ecology,* 88**,** 1322-7.

Bogaert, D., De Groot, R. & Hermans, P. W. 2004a. Streptococcus Pneumoniae Colonisation: The Key To Pneumococcal Disease. *The Lancet Infectious Diseases,* 4**,** 144-54.

Bogaert, D., Engelen, M. N., Timmers-Reker, A. J., Elzenaar, K. P., Peerbooms, P. G., Coutinho, R. A., De Groot, R. & Hermans, P. W. 2001. Pneumococcal Carriage In Children In The Netherlands: A Molecular Epidemiological Study. *J Clin Microbiol,* 39**,** 3316-20.

Bogaert, D., Hermans, P. W., Adrian, P. V., Rumke, H. C. & De Groot, R. 2004b. Pneumococcal Vaccines: An Update On Current Strategies. *Vaccine,* 22**,** 2209-20.

Bogaert, D., Hermans, P. W., Boelens, H., Sluijter, M., Luijendijk, A., Rumke, H. C., Koppen, S., Van Belkum, A., De Groot, R. & Verbrugh, H. A. 2005. Epidemiology Of Nasopharyngeal Carriage Of Neisseria Meningitidis In Healthy Dutch Children. *Clin Infect Dis,* 40**,** 899-902.

Bogaert, D., Keijser, B., Huse, S., Rossen, J., Veenhoven, R., Van Gils, E., Bruin, J., Montijn, R., Bonten, M. & Sanders, E. 2011. Variability And Diversity Of Nasopharyngeal Microbiota In Children: A Metagenomic Analysis. *Plos One,* 6**,** E17035.

Bogaert, D., Thompson, C. M., Trzcinski, K., Malley, R. & Lipsitch, M. 2010. The Role Of Complement In Innate And Adaptive Immunity To Pneumococcal Colonization And Sepsis In A Murine Model. *Vaccine,* 28**,** 681-5.

Bogaert, D., Van Belkum, A., Sluijter, M., Luijendijk, A., De Groot, R., Rumke, H. C., Verbrugh, H. A. & Hermans, P. W. 2004c. Colonisation By Streptococcus

Pneumoniae And Staphylococcus Aureus In Healthy Children. *Lancet,* 363**,** 1871-2.

Bogaert, D., Weinberger, D., Thompson, C., Lipsitch, M. & Malley, R. 2009. Impaired Innate And Adaptive Immunity To Streptococcus Pneumoniae And Its Effect On Colonization In An Infant Mouse Model. *Infect Immun,* 77**,** 1613-22.

Boivin-Jahns, V., Bianchi, A., Ruimy, R., Garcin, J., Daumas, S. & Christen, R. 1995. Comparison Of Phenotypical And Molecular Methods For The Identification Of Bacterial Strains Isolated From A Deep Subsurface Environment. *Appl Environ Microbiol,* 61**,** 4140.

Bosch, A. A., Biesbroek, G., Trzcinski, K., Sanders, E. A. & Bogaert, D. 2013. Viral And Bacterial Interactions In The Upper Respiratory Tract. *Plos Pathogens,* 9**,** E1003057.

Bowers, E. F. & Jeffries, L. R. 1955. Optochin In The Identification Of Str. Pneumoniae. *Journal Of Clinical Pathology,* 8**,** 58-60.

Bratcher, P. E., Kim, K. H., Kang, J. H., Hong, J. Y. & Nahm, M. H. 2010. Identification Of Natural Pneumococcal Isolates Expressing Serotype 6d By Genetic, Biochemical And Serological Characterization. *Microbiology,* 156**,** 555-60.

Bray, J. R. & Curtis, J. T. 1957. An Ordination Of Upland Forest Communities Of Southern Winsconsin. *Ecological Monographs,* 27**,** 325-349.

Brito, D. A., Ramirez, M. & De Lencastre, H. 2003. Serotyping Streptococcus Pneumoniae By Multiplex Pcr. *J Clin Microbiol,* 41**,** 2378-84.

Brown, J. S., Gilliland, S. M. & Holden, D. W. 2001a. A Streptococcus Pneumoniae Pathogenicity Island Encoding An Abc Transporter Involved In Iron Uptake And Virulence. *Mol Microbiol,* 40**,** 572-85.

Brown, J. S., Ogunniyi, A. D., Woodrow, M. C., Holden, D. W. & Paton, J. C. 2001b. Immunization With Components Of Two Iron Uptake Abc Transporters Protects Mice Against Systemic Streptococcus Pneumoniae Infection. *Infect Immun,* 69**,** 6702-6.

Brueggemann, A. B., Griffiths, D. T., Meats, E., Peto, T., Crook, D. W. & Spratt, B. G. 2003. Clonal Relationships Between Invasive And Carriage Streptococcus Pneumoniae And Serotype- And Clone-Specific Differences In Invasive Disease Potential. *J Infect Dis,* 187**,** 1424-32.

Brueggemann, A. B., Pai, R., Crook, D. W. & Beall, B. 2007a. Vaccine Escape Recombinants Emerge After Pneumococcal Vaccination In The United States. *Plos Pathogens,* 3**,** E168.

Brueggemann, A. B., Pai, R., Crook, D. W. & Beall, B. 2007b. Vaccine Escape Recombinants Emerge After Pneumococcal Vaccination In The United States. *Plos Pathog,* 3**,** E168.

Brugger, S. D., Frey, P., Aebi, S., Hinds, J. & Muhlemann, K. 2010. Multiple Colonization With S. Pneumoniae Before And After Introduction Of The Seven-Valent Conjugated Pneumococcal Polysaccharide Vaccine. *Plos One,* 5**,** E11638.

Brugger, S. D., Hathaway, L. J. & Muhlemann, K. 2009. Detection Of Streptococcus Pneumoniae Strain Cocolonization In The Nasopharynx. *J Clin Microbiol,* 47**,** 1750-6.

Bui, N. K., Gray, J., Schwarz, H., Schumann, P., Blanot, D. & Vollmer, W. 2009. The Peptidoglycan Sacculus Of Myxococcus Xanthus Has Unusual Structural Features And Is Degraded During Glycerol-Induced Myxospore Development. *Journal Of Bacteriology,* 191**,** 494-505.

Calix, J. J., Dagan, R., Pelton, S. I., Porat, N. & Nahm, M. H. 2012a. Differential Occurrence Of Streptococcus Pneumoniae Serotype 11e Between Asymptomatic Carriage And Invasive Pneumococcal Disease Isolates Reflects A Unique Model Of Pathogen Microevolution. *Clin Infect Dis,* 54**,** 794-9.

Calix, J. J., Porambo, R. J., Brady, A. M., Larson, T. R., Yother, J., Abeygunwardana, C. & Nahm, M. H. 2012b. Biochemical, Genetic, And Serological Characterization Of Two Capsule Subtypes Among Streptococcus Pneumoniae Serotype 20 Strains: Discovery Of A New Pneumococcal Serotype. *J Biol Chem,* 287**,** 27885-94.

Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M. L. & Brussow, H. 2003. Phage As Agents Of Lateral Gene Transfer. *Current Opinion In Microbiology,* 6**,** 417-24.

Carrol, E. D., Guiver, M., Nkhoma, S., Mankhambo, L. A., Marsh, J., Balmer, P., Banda, D. L., Jeffers, G., White, S. A., Molyneux, E. M., Molyneux, M. E., Smyth, R. L. & Hart, C. A. 2007. High Pneumococcal Dna Loads Are

Associated With Mortality In Malawian Children With Invasive Pneumococcal Disease. *Pediatr Infect Dis J,* 26**,** 416-22.

Cartee, R. T., Forsee, W. T., Schutzbach, J. S. & Yother, J. 2000. Mechanism Of Type 3 Capsular Polysaccharide Synthesis In Streptococcus Pneumoniae. *The Journal Of Biological Chemistry,* 275**,** 3907-14.

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M. A., Barrell, B. G. & Parkhill, J. 2005. Act: The Artemis Comparison Tool. *Bioinformatics,* 21**,** 3422-3.

Cassol, E., Malfeld, S., Mahasha, P., Bond, R., Slavik, T., Seebregts, C., Poli, G., Cassol, S., Van Der Merwe, S. W. & Rossouw, T. 2013. Impaired Cd4+ T-Cell Restoration In The Small Versus Large Intestine Of Hiv-1-Positive South Africans Receiving Combination Antiretroviral Therapy. *The Journal Of Infectious Diseases,* 208**,** 1113-22.

Chaban, B., Albert, A., Links, M. G., Gardy, J., Tang, P. & Hill, J. E. 2013. Characterization Of The Upper Respiratory Tract Microbiomes Of Patients With Pandemic H1n1 Influenza. *Plos One,* 8**,** E69559.

Chakravorty, S., Helb, D., Burday, M., Connell, N. & Alland, D. 2007. A Detailed Analysis Of 16s Ribosomal Rna Gene Segments For The Diagnosis Of Pathogenic Bacteria. *Journal Of Microbiological Methods,* 69**,** 330-9.

Chen, I., Christie, P. J. & Dubnau, D. 2005. The Ins And Outs Of Dna Transfer In Bacteria. *Science,* 310**,** 1456-60.

Chewapreecha, C., Harris, S. R., Croucher, N. J., Turner, C., Marttinen, P., Cheng, L., Pessia, A., Aanensen, D. M., Mather, A. E., Page, A. J., Salter, S. J., Harris, D., Nosten, F., Goldblatt, D., Corander, J., Parkhill, J., Turner, P. & Bentley, S. D. 2014. Dense Genomic Sampling Identifies Highways Of Pneumococcal Recombination. *Nat Genet,* 46**,** 305-9.

Cho, I. & Blaser, M. J. 2012. The Human Microbiome: At The Interface Of Health And Disease. *Nat Rev Genet,* 13**,** 260-70.

Chou, H. H. & Holmes, M. H. 2001. Dna Sequence Quality Trimming And Vector Removal. *Bioinformatics,* 17**,** 1093-104.

Chu, Y. W., Leung, C. M., Houang, E. T., Ng, K. C., Leung, C. B., Leung, H. Y. & Cheng, A. F. 1999. Skin Carriage Of Acinetobacters In Hong Kong. *J Clin Microbiol,* 37**,** 2962-7.

Claesson, M. J., Wang, Q., O'sullivan, O., Greene-Diniz, R., Cole, J. R., Ross, R. P. & O'toole, P. W. 2010. Comparison Of Two Next-Generation Sequencing Technologies For Resolving Highly Complex Microbiota Composition Using Tandem Variable 16s Rrna Gene Regions. *Nucleic Acids Research,* 38**,** E200.

Clarridge, J. E., 3rd 2004. Impact Of 16s Rrna Gene Sequence Analysis For Identification Of Bacteria On Clinical Microbiology And Infectious Diseases. *Clinical Microbiology Reviews,* 17**,** 840-62, Table Of Contents.

Claverys, J. P., Martin, B. & Polard, P. 2009. The Genetic Transformation Machinery: Composition, Localization, And Mechanism. *Fems Microbiology Reviews,* 33**,** 643-56.

Clemente, J. C., Ursell, L. K., Parfrey, L. W. & Knight, R. 2012. The Impact Of The Gut Microbiota On Human Health: An Integrative View. *Cell,* 148**,** 1258-70.

Cobey, S. & Lipsitch, M. 2012. Niche And Neutral Effects Of Acquired Immunity Permit Coexistence Of Pneumococcal Serotypes. *Science,* 335**,** 1376-80.

Cochetti, I., Vecchi, M., Mingoia, M., Tili, E., Catania, M. R., Manzin, A., Varaldo, P. E. & Montanari, M. P. 2005. Molecular Characterization Of Pneumococci With Efflux-Mediated Erythromycin Resistance And Identification Of A Novel Mef Gene Subclass, Mef(I). *Antimicrob Agents Chemother,* 49**,** 4999-5006.

Cohen, R., Levy, C., Bonnet, E., Thollot, F., Boucherat, M., Fritzell, B., Derkx, V., Bingen, E. & Varon, E. 2011. Risk Factors For Serotype 19a Carriage After Introduction Of 7-Valent Pneumococcal Vaccination. *Bmc Infectious Diseases,* 11**,** 95.

Coles, C. L., Kanungo, R., Rahmathullah, L., Thulasiraj, R. D., Katz, J., Santosham, M. & Tielsch, J. M. 2001. Pneumococcal Nasopharyngeal Colonization In Young South Indian Infants. *Pediatr Infect Dis J,* 20**,** 289-95.

Colwell, R. K., Chao, A., Gotelli, N. J., Lin, S., Mao, C. X., Chazdon, R. L. & Longino, J. T. 2012. Models And Estimators Linking Individual-Based And Sample-Based Rarefaction, Extrapolation And Comparison Of Assemblages. *Journal Of Plant Ecology,* 5**,** 3-21.

Consortium., T. H. M. P. 2012. Structure, Function And Diversity Of The Healthy Human Microbiome. *Nature,* 486**,** 207-14.

Coonrod, J. D., Varble, R. & Yoneda, K. 1991. Mechanism Of Killing Of Pneumococci By Lysozyme. *The Journal Of Infectious Diseases,* 164**,** 527-32.

Cornick, J. E., Everett, D. B., Broughton, C., Denis, B. B., Banda, D. L., Carrol, E. D. & Parry, C. M. 2011. Invasive Streptococcus Pneumoniae In Children, Malawi, 2004-2006. *Emerg Infect Dis,* 17**,** 1107-9.

Cornick, J. E., Harris, S. R., Parry, C. M., Moore, M. J., Jassi, C., Kamng'ona, A., Kulohoma, B., Heyderman, R. S., Bentley, S. D. & Everett, D. B. 2013. Genomic Identification Of A Novel Co-Trimoxazole Resistance Genotype And Its Prevalence Amongst Streptococcus Pneumoniae In Malawi. *The Journal Of Antimicrobial Chemotherapy,* 69**,** 368-74.

Costa, M. C., Arroyo, L. G., Allen-Vercoe, E., Stampfli, H. R., Kim, P. T., Sturgeon, A. & Weese, J. S. 2012. Comparison Of The Fecal Microbiota Of Healthy Horses And Horses With Colitis By High Throughput Sequencing Of The V3-V5 Region Of The 16s Rrna Gene. *Plos One,* 7**,** E41484.

Croucher, N. J., Finkelstein, J. A., Pelton, S. I., Mitchell, P. K., Lee, G. M., Parkhill, J., Bentley, S. D., Hanage, W. P. & Lipsitch, M. 2013. Population Genomics Of Post-Vaccine Changes In Pneumococcal Epidemiology. *Nature Genetics,* 45**,** 656-63.

Croucher, N. J., Harris, S. R., Fraser, C., Quail, M. A., Burton, J., Van Der Linden, M., Mcgee, L., Von Gottberg, A., Song, J. H., Ko, K. S., Pichon, B., Baker, S., Parry, C. M., Lambertsen, L. M., Shahinas, D., Pillai, D. R., Mitchell, T. J., Dougan, G., Tomasz, A., Klugman, K. P., Parkhill, J., Hanage, W. P. &

Bentley, S. D. 2011. Rapid Pneumococcal Evolution In Response To Clinical Interventions. *Science,* 331**,** 430-4.

Croucher, N. J., Walker, D., Romero, P., Lennard, N., Paterson, G. K., Bason, N. C., Mitchell, A. M., Quail, M. A., Andrew, P. W., Parkhill, J., Bentley, S. D. & Mitchell, T. J. 2009. Role Of Conjugative Elements In The Evolution Of The Multidrug-Resistant Pandemic Clone Streptococcus Pneumoniaespain23f St81. *J Bacteriol,* 191**,** 1480-9.

Cundell, D. R., Gerard, N. P., Gerard, C., Idanpaan-Heikkila, I. & Tuomanen, E. I. 1995a. Streptococcus Pneumoniae Anchor To Activated Human Cells By The Receptor For Platelet-Activating Factor. *Nature,* 377**,** 435-8.

Cundell, D. R., Weiser, J. N., Shen, J., Young, A. & Tuomanen, E. I. 1995b. Relationship Between Colonial Morphology And Adherence Of Streptococcus Pneumoniae. *Infection And Immunity,* 63**,** 757-61.

Cvitkovitch, D. G., Li, Y. H. & Ellen, R. P. 2003. Quorum Sensing And Biofilm Formation In Streptococcal Infections. *J Clin Invest,* 112**,** 1626-32.

Da Gloria Carvalho, M., Pimenta, F. C., Jackson, D., Roundtree, A., Ahmad, Y., Millar, E. V., O'brien, K. L., Whitney, C. G., Cohen, A. L. & Beall, B. W. 2010. Revisiting Pneumococcal Carriage By Use Of Broth Enrichment And Pcr Techniques For Enhanced Detection Of Carriage And Serotypes. *Journal Of Clinical Microbiology,* 48**,** 1611-8.

Dagan, R. 2009. New Insights On Pneumococcal Disease: What We Have Learned Over The Past Decade. *Vaccine,* 27 Suppl 3**,** C3-5.

Dagan, R., Givon-Lavi, N., Zamir, O., Sikuler-Cohen, M., Guy, L., Janco, J., Yagupsky, P. & Fraser, D. 2002. Reduction Of Nasopharyngeal Carriage Of Streptococcus Pneumoniae After Administration Of A 9-Valent Pneumococcal Conjugate Vaccine To Toddlers Attending Day Care Centers. *J Infect Dis,* 185**,** 927-36.

Dagan, R., Leibovitz, E., Greenberg, D., Bakaletz, L. & Givon-Lavi, N. 2013. Mixed Pneumococcal-Nontypeable Haemophilus Influenzae Otitis Media Is A Distinct Clinical Entity With Unique Epidemiologic Characteristics And Pneumococcal Serotype Distribution. *J Infect Dis,* 208**,** 1152-60.

Dagan, R., Muallem, M., Melamed, R., Leroy, O. & Yagupsky, P. 1997. Reduction Of Pneumococcal Nasopharyngeal Carriage In Early Infancy After Immunization With Tetravalent Pneumococcal Vaccines Conjugated To Either Tetanus Toxoid Or Diphtheria Toxoid. *Pediatr Infect Dis J,* 16**,** 1060-4.

Darrieux, M., Goulart, C., Briles, D. & Leite, L. C. 2013. Current Status And Perspectives On Protein-Based Pneumococcal Vaccines. *Crit Rev Microbiol*.

Dave, S., Carmicle, S., Hammerschmidt, S., Pangburn, M. K. & Mcdaniel, L. S. 2004a. Dual Roles Of Pspc, A Surface Protein Of Streptococcus Pneumoniae, In Binding Human Secretory Iga And Factor H. *Journal Of Immunology,* 173**,** 471-7.

Dave, S., Carmicle, S., Hammerschmidt, S., Pangburn, M. K. & Mcdaniel, L. S. 2004b. Dual Roles Of Pspc, A Surface Protein Of Streptococcus Pneumoniae, In Binding Human Secretory Iga And Factor H. *J Immunol,* 173**,** 471-7.

Dave, S., Pangburn, M. K., Pruitt, C. & Mcdaniel, L. S. 2004c. Interaction Of Human Factor H With Pspc Of Streptococcus Pneumoniae. *Indian J Med Res,* 119 Suppl**,** 66-73.

Davis, K. M., Akinbi, H. T., Standish, A. J. & Weiser, J. N. 2008. Resistance To Mucosal Lysozyme Compensates For The Fitness Deficit Of Peptidoglycan Modifications By Streptococcus Pneumoniae. *Plos Pathogens,* 4**,** E1000241.

De Clerck, E. & De Vos, P. 2004. Genotypic Diversity Among Bacillus Licheniformis Strains From Various Sources. *Fems Microbiol Lett,* 231**,** 91-8.

Dias, C. A., Teixeira, L. M., Carvalho Mda, G. & Beall, B. 2007. Sequential Multiplex Pcr For Determining Capsular Serotypes Of Pneumococci Recovered From Brazilian Children. *Journal Of Medical Microbiology,* 56**,** 1185-8.

Diavatopoulos, D. A., Short, K. R., Price, J. T., Wilksch, J. J., Brown, L. E., Briles, D. E., Strugnell, R. A. & Wijburg, O. L. 2010. Influenza A Virus Facilitates Streptococcus Pneumoniae Transmission And Disease. *Faseb Journal : Official Publication Of The Federation Of American Societies For Experimental Biology,* 24**,** 1789-98.

Ding, F., Tang, P., Hsu, M. H., Cui, P., Hu, S., Yu, J. & Chiu, C. H. 2009. Genome Evolution Driven By Host Adaptations Results In A More Virulent And Antimicrobial-Resistant Streptococcus Pneumoniae Serotype 14. *Bmc Genomics,* 10**,** 158.

Dintilhac, A., Alloing, G., Granadel, C. & Claverys, J. P. 1997. Competence And Virulence Of Streptococcus Pneumoniae: Adc And Psaa Mutants Exhibit A Requirement For Zn And Mn Resulting From Inactivation Of Putative Abc Metal Permeases. *Mol Microbiol,* 25**,** 727-39.

Domenech, M., Ramos-Sevillano, E., Garcia, E., Moscoso, M. & Yuste, J. 2013. Biofilm Formation Avoids Complement Immunity And Phagocytosis Of Streptococcus Pneumoniae. *Infect Immun,* 81**,** 2606-15.

Domingo, P., Suarez-Lozano, I., Torres, F., Pomar, V., Ribera, E., Galindo, M. J., Cosin, J., Garcia-Alcalde, M. L., Vidal, F., Lopez-Aldeguer, J., Roca, B., Gonzalez, J., Lozano, F. & Garrido, M. 2009. Bacterial Meningitis In Hiv-1-Infected Patients In The Era Of Highly Active Antiretroviral Therapy. *J Acquir Immune Defic Syndr,* 51**,** 582-7.

Donkor, E. S., Bishop, C. J., Antonio, M., Wren, B. & Hanage, W. P. 2011. High Levels Of Recombination Among Streptococcus Pneumoniae Isolates From The Gambia. *Mbio,* 2**,** E00040-11.

Doolittle, R. F. 2002. Biodiversity: Microbial Genomes Multiply. *Nature,* 416**,** 697-700.

Dorrell, N., Champion, O. L. & Wren, B. W. 2002. Application Of Dna Microarrays For Comparative And Evolutionary Genomics. *In:* Wren, B. W. & Dorrell, N. (Eds.) *Functional Microbial Genomics.* Amsterdam: Academic Press. pp. 121- 136

Drancourt, M., Bollet, C., Carlioz, A., Martelin, R., Gayral, J. P. & Raoult, D. 2000. 16s Ribosomal Dna Sequence Analysis Of A Large Collection Of

Environmental And Clinical Unidentifiable Bacterial Isolates. *Journal Of Clinical Microbiology,* 38**,** 3623-30.

Dressman, D., Yan, H., Traverso, G., Kinzler, K. W. & Vogelstein, B. 2003. Transforming Single Dna Molecules Into Fluorescent Magnetic Particles For Detection And Enumeration Of Genetic Variations. *Proceedings Of The National Academy Of Sciences Of The United States Of America,* 100**,** 8817-22.

Du Clos, T. W. & Mold, C. 2004. C-Reactive Protein: An Activator Of Innate Immunity And A Modulator Of Adaptive Immunity. *Immunol Res,* 30**,** 261-77.

Dunne, E. M., Smith-Vaughan, H. C., Robins-Browne, R. M., Mulholland, E. K. & Satzke, C. 2013. Nasopharyngeal Microbial Interactions In The Era Of Pneumococcal Conjugate Vaccination. *Vaccine,* 31**,** 2333-42.

Edelstein, P. H. 2004. Pneumococcal Resistance To Macrolides, Lincosamides, Ketolides, And Streptogramin B Agents: Molecular Mechanisms And Resistance Phenotypes. *Clin Infect Dis,* 38 Suppl 4**,** S322-7.

Edgar, R. C. 2004. Muscle: Multiple Sequence Alignment With High Accuracy And High Throughput. *Nucleic Acids Research,* 32**,** 1792-7.

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. 2011. Uchime Improves Sensitivity And Speed Of Chimera Detection. *Bioinformatics,* 27**,** 2194-200.

Efron, B., Halloran, E. & Holmes, S. 1996. Bootstrap Confidence Levels For Phylogenetic Trees. *Proceedings Of The National Academy Of Sciences Of The United States Of America,* 93**,** 13429-34.

Elberse, K., Witteveen, S., Van Der Heide, H., Van De Pol, I., Schot, C., Van Der Ende, A., Berbers, G. & Schouls, L. 2011. Sequence Diversity Within The Capsular Genes Of Streptococcus Pneumoniae Serogroup 6 And 19. *Plos One,* 6**,** E25018.

Elm, C., Braathen, R., Bergmann, S., Frank, R., Vaerman, J. P., Kaetzel, C. S., Chhatwal, G. S., Johansen, F. E. & Hammerschmidt, S. 2004. Ectodomains 3 And 4 Of Human Polymeric Immunoglobulin Receptor (Hpigr) Mediate Invasion Of Streptococcus Pneumoniae Into The Epithelium. *The Journal Of Biological Chemistry,* 279**,** 6296-304.

Enright, M. C. & Spratt, B. G. 1998. A Multilocus Sequence Typing Scheme For Streptococcus Pneumoniae: Identification Of Clones Associated With Serious Invasive Disease. *Microbiology,* 144 (Pt 11)**,** 3049-60.

Ercibengoa, M., Arostegi, N., Marimon, J. M., Alonso, M. & Perez-Trallero, E. 2012. Dynamics Of Pneumococcal Nasopharyngeal Carriage In Healthy Children Attending A Day Care Center In Northern Spain. Influence Of Detection Techniques On The Results. *Bmc Infectious Diseases,* 12**,** 69.

Eskola, J. & Anttila, M. 1999. Pneumococcal Conjugate Vaccines. *Pediatr Infect Dis J,* 18**,** 543-51.

Esposito, S., Tansey, S., Thompson, A., Razmpour, A., Liang, J., Jones, T. R., Ferrera, G., Maida, A., Bona, G., Sabatini, C., Pugni, L., Emini, E. A.,

Gruber, W. C., Scott, D. A. & Principi, N. 2010. Safety And Immunogenicity Of A 13-Valent Pneumococcal Conjugate Vaccine Compared To Those Of A 7-Valent Pneumococcal Conjugate Vaccine Given As A Three-Dose Series With Routine Vaccines In Healthy Infants And Toddlers. *Clin Vaccine Immunol,* 17**,** 1017-26.

Esteva, C., Selva, L., De Sevilla, M. F., Garcia-Garcia, J. J., Pallares, R. & Munoz-Almagro, C. 2011. Streptococcus Pneumoniae Serotype 1 Causing Invasive Disease Among Children In Barcelona Over A 20-Year Period (1989-2008). *Clin Microbiol Infect,* 17**,** 1441-4.

Everett, D. B., Cornick, J., Denis, B., Chewapreecha, C., Croucher, N., Harris, S., Parkhill, J., Gordon, S., Carrol, E. D., French, N., Heyderman, R. S. & Bentley, S. D. 2012. Genetic Characterisation Of Malawian Pneumococci Prior To The Roll-Out Of The Pcv13 Vaccine Using A High-Throughput Whole Genome Sequencing Approach. *Plos One,* 7**,** E44250.

Everett, D. B., Mukaka, M., Denis, B., Gordon, S. B., Carrol, E. D., Van Oosterhout, J. J., Molyneux, E. M., Molyneux, M., French, N. & Heyderman, R. S. 2011. Ten Years Of Surveillance For Invasive Streptococcus Pneumoniae During The Era Of Antiretroviral Scale-Up And Cotrimoxazole Prophylaxis In Malawi. *Plos One,* 6**,** E17765.

Facklam, R. 2002. What Happened To The Streptococci: Overview Of Taxonomic And Nomenclature Changes. *Clin Microbiol Rev,* 15**,** 613-30.

Faden, H., Duffy, L., Wasielewski, R., Wolf, J., Krystofik, D. & Tung, Y. 1997. Relationship Between Nasopharyngeal Colonization And The Development

Of Otitis Media In Children. Tonawanda/Williamsville Pediatrics. *J Infect Dis,* 175**,** 1440-5.

Faden, H., Duffy, L., Williams, A., Krystofik, D. A. & Wolf, J. 1996. Epidemiology Of Nasopharyngeal Colonization With Nontypeable Haemophilus Influenzae In The First Two Years Of Life. *Acta Otolaryngol Suppl,* 523**,** 128-9.

Farr, B. M., Bartlett, C. L., Wadsworth, J. & Miller, D. L. 2000a. Risk Factors For Community-Acquired Pneumonia Diagnosed Upon Hospital Admission. British Thoracic Society Pneumonia Study Group. *Respir Med,* 94**,** 954-63.

Farr, B. M., Woodhead, M. A., Macfarlane, J. T., Bartlett, C. L., Mccraken, J. S., Wadsworth, J. & Miller, D. L. 2000b. Risk Factors For Community-Acquired Pneumonia Diagnosed By General Practitioners In The Community. *Respir Med,* 94**,** 422-7.

Fasching, C. E., Grossman, T., Corthesy, B., Plaut, A. G., Weiser, J. N. & Janoff, E. N. 2007. Impact Of The Molecular Form Of Immunoglobulin A On Functional Activity In Defense Against Streptococcus Pneumoniae. *Infection And Immunity,* 75**,** 1801-10.

Fedurco, M., Romieu, A., Williams, S., Lawrence, I. & Turcatti, G. 2006. Bta, A Novel Reagent For Dna Attachment On Glass And Efficient Generation Of Solid-Phase Amplified Dna Colonies. *Nucleic Acids Research,* 34**,** E22.

Feikin, D. R., Davis, M., Nwanyanwu, O. C., Kazembe, P. N., Barat, L. M., Wasas, A., Bloland, P. B., Ziba, C., Capper, T., Huebner, R. E., Schwartz, B., Klugman, K. P. & Dowell, S. F. 2003. Antibiotic Resistance And Serotype

Distribution Of Streptococcus Pneumoniae Colonizing Rural Malawian Children. *Pediatr Infect Dis J,* 22**,** 564-7.

Feikin, D. R., Klugman, K. P., Facklam, R. R., Zell, E. R., Schuchat, A., Whitney, C. G. & Network, A. B. C. S. E. I. P. 2005. Increased Prevalence Of Pediatric Pneumococcal Serotypes In Elderly Adults. *Clin Infect Dis,* 41**,** 481-7.

Feil, E. J., Enright, M. C. & Spratt, B. G. 2000. Estimating The Relative Contributions Of Mutation And Recombination To Clonal Diversification: A Comparison Between Neisseria Meningitidis And Streptococcus Pneumoniae. *Res Microbiol,* 151**,** 465-9.

Feldman, C., Anderson, R., Cockeran, R., Mitchell, T., Cole, P. & Wilson, R. 2002. The Effects Of Pneumolysin And Hydrogen Peroxide, Alone And In Combination, On Human Ciliated Epithelium In Vitro. *Respiratory Medicine,* 96**,** 580-5.

Feldman, C., Mitchell, T. J., Andrew, P. W., Boulnois, G. J., Read, R. C., Todd, H. C., Cole, P. J. & Wilson, R. 1990. The Effect Of Streptococcus Pneumoniae Pneumolysin On Human Respiratory Epithelium In Vitro. *Microbial Pathogenesis,* 9**,** 275-84.

Fenart, S., Chabi, M., Gallina, S., Huis, R., Neutelings, G., Riviere, N., Thomasset, B., Hawkins, S. & Lucau-Danila, A. 2013. Intra-Platform Comparison Of 25-Mer And 60-Mer Oligonucleotide Nimblegen Dna Microarrays. *Bmc Research Notes,* 6**,** 43.

Ferreira, D. M., Darrieux, M., Silva, D. A., Leite, L. C., Ferreira, J. M., Jr., Ho, P. L., Miyaji, E. N. & Oliveira, M. L. 2009. Characterization Of Protective Mucosal

And Systemic Immune Responses Elicited By Pneumococcal Surface Protein Pspa And Pspc Nasal Vaccines Against A Respiratory Pneumococcal Challenge In Mice. *Clin Vaccine Immunol,* 16**,** 636-45.

Ferreira, D. M., Neill, D. R., Bangert, M., Gritzfeld, J. F., Green, N., Wright, A. K., Pennington, S. H., Bricio-Moreno, L., Moreno, A. T., Miyaji, E. N., Wright, A. D., Collins, A. M., Goldblatt, D., Kadioglu, A. & Gordon, S. B. 2013. Controlled Human Infection And Rechallenge With Streptococcus Pneumoniae Reveals The Protective Efficacy Of Carriage In Healthy Adults. *American Journal Of Respiratory And Critical Care Medicine,* 187**,** 855-64.

Finn, A., Zhang, Q., Seymour, L., Fasching, C., Pettitt, E. & Janoff, E. N. 2002. Induction Of Functional Secretory Iga Responses In Breast Milk, By Pneumococcal Capsular Polysaccharides. *J Infect Dis,* 186**,** 1422-9.

Flasche, S., Slack, M. & Miller, E. 2011a. Long Term Trends Introduce A Potential Bias When Evaluating The Impact Of The Pneumococcal Conjugate Vaccination Programme In England And Wales. *Euro Surveill,* 16**,** 19868.

Flasche, S., Van Hoek, A. J., Sheasby, E., Waight, P., Andrews, N., Sheppard, C., George, R. & Miller, E. 2011b. Effect Of Pneumococcal Conjugate Vaccination On Serotype-Specific Carriage And Invasive Disease In England: A Cross-Sectional Study. *Plos Medicine,* 8**,** E1001017.

Fournier, P. E. & Richet, H. 2006. The Epidemiology And Control Of Acinetobacter Baumannii In Health Care Facilities. *Clin Infect Dis,* 42**,** 692-9.

Fox, G. E., Wisotzkey, J. D. & Jurtshuk, P., Jr. 1992. How Close Is Close: 16s Rrna Sequence Identity May Not Be Sufficient To Guarantee Species Identity. *Int J Syst Bacteriol,* 42**,** 166-70.

Frank, D. N., Feazel, L. M., Bessesen, M. T., Price, C. S., Janoff, E. N. & Pace, N. R. 2010. The Human Nasal Microbiota And Staphylococcus Aureus Carriage. *Plos One,* 5**,** E10598.

French, N., Gilks, C. F., Mujugira, A., Fasching, C., O'brien, J. & Janoff, E. N. 1998. Pneumococcal Vaccination In Hiv-1-Infected Adults In Uganda: Humoral Response And Two Vaccine Failures. *Aids,* 12**,** 1683-9.

French, N., Gordon, S. B., Mwalukomo, T., White, S. A., Mwafulirwa, G., Longwe, H., Mwaiponya, M., Zijlstra, E. E., Molyneux, M. E. & Gilks, C. F. 2010. A Trial Of A 7-Valent Pneumococcal Conjugate Vaccine In Hiv-Infected Adults. *N Engl J Med,* 362**,** 812-22.

French, N., Nakiyingi, J., Carpenter, L. M., Lugada, E., Watera, C., Moi, K., Moore, M., Antvelink, D., Mulder, D., Janoff, E. N., Whitworth, J. & Gilks, C. F. 2000. 23-Valent Pneumococcal Polysaccharide Vaccine In Hiv-1-Infected Ugandan Adults: Double-Blind, Randomised And Placebo Controlled Trial. *Lancet,* 355**,** 2106-11.

Ganz, T. 2003. Defensins: Antimicrobial Peptides Of Innate Immunity. *Nature Reviews. Immunology,* 3**,** 710-20.

Garcia-Rodriguez, J. A. & Fresnadillo Martinez, M. J. 2002. Dynamics Of Nasopharyngeal Colonization By Potential Respiratory Pathogens. *J Antimicrob Chemother,* 50 Suppl S2**,** 59-73.

Gardam, M. A. & Miller, M. A. 1998. Optochin Revisited: Defining The Optimal Type Of Blood Agar For Presumptive Identification Of Streptococcus Pneumoniae. *Journal Of Clinical Microbiology,* 36**,** 833-4.

Gebo, K. A., Gallant, J. E., Keruly, J. C. & Moore, R. D. 2004. Absolute Cd4 Vs. Cd4 Percentage For Predicting The Risk Of Opportunistic Illness In Hiv Infection. *J Acquir Immune Defic Syndr,* 36**,** 1028-33.

Gella, J. F., Serra, J. & Gener, J. 1991. Latex Agglutination Procedure In Immunodiagnosis. *Pure & Appl. Chem,* 63**,** 1131-1134.

Gibson, L. F. & Khoury, J. T. 1986. Storage And Survival Of Bacteria By Ultra-Freeze. *Letters In Applied Microbiology,* 3**,** 127-129.

Gilks, C. F., Ojoo, S. A., Ojoo, J. C., Brindle, R. J., Paul, J., Batchelor, B. I., Kimari, J. N., Newnham, R., Bwayo, J., Plummer, F. A. & Et Al. 1996. Invasive Pneumococcal Disease In A Cohort Of Predominantly Hiv-1 Infected Female Sex-Workers In Nairobi, Kenya. *Lancet,* 347**,** 718-23.

Gill, C. J., Mwanakasale, V., Fox, M. P., Chilengi, R., Tembo, M., Nsofwa, M., Chalwe, V., Mwananyanda, L., Mukwamataba, D., Malilwe, B., Champo, D., Macleod, W. B., Thea, D. M. & Hamer, D. H. 2008a. Effect Of Presumptive Co-Trimoxazole Prophylaxis On Pneumococcal Colonization Rates, Seroepidemiology And Antibiotic Resistance In Zambian Infants: A Longitudinal Cohort Study. *Bull World Health Organ,* 86**,** 929-38.

Gill, C. J., Mwanakasale, V., Fox, M. P., Chilengi, R., Tembo, M., Nsofwa, M., Chalwe, V., Mwananyanda, L., Mukwamataba, D., Malilwe, B., Champo, D., Macleod, W. B., Thea, D. M. & Hamer, D. H. 2008b. Impact Of Human

Immunodeficiency Virus Infection On Streptococcus Pneumoniae Colonization And Seroepidemiology Among Zambian Women. *J Infect Dis,* 197**,** 1000-5.

Gillespie, S. H. & Balakrishnan, I. 2000. Pathogenesis Of Pneumococcal Infection. *J Med Microbiol,* 49**,** 1057-67.

Givon-Lavi, N., Fraser, D., Porat, N. & Dagan, R. 2002. Spread Of Streptococcus Pneumoniae And Antibiotic-Resistant S. Pneumoniae From Day-Care Center Attendees To Their Younger Siblings. *J Infect Dis,* 186**,** 1608-14.

Glennie, S. J., Banda, D., Gould, K., Hinds, J., Kamng'ona, A., Everett, D. D., Williams, N. A. & Heyderman, R. S. 2012a. Defective Pneumococcal-Specific Th1 Responses In Hiv-Infected Adults Precedes A Loss Of Control Of Pneumococcal Colonization. *Clin Infect Dis,* 56**,** 291-9.

Glennie, S. J., Banda, D., Mulwafu, W., Nkhata, R., Williams, N. A. & Heyderman, R. S. 2012b. Regulation Of Naturally Acquired Mucosal Immunity To Streptococcus Pneumoniae In Healthy Malawian Adults And Children. *Plos One,* 7**,** E51425.

Glennie, S. J., Williams, N. A. & Heyderman, R. S. 2010. Mucosal Immunity In Resource-Limited Setting: Is The Battle Ground Different? *Trends Microbiol,* 18**,** 487-93.

Glover, D. T., Hollingshead, S. K. & Briles, D. E. 2008. Streptococcus Pneumoniae Surface Protein Pcpa Elicits Protection Against Lung Infection And Fatal Sepsis. *Infect Immun,* 76**,** 2767-76.

Godfroid, F., Hermand, P., Verlant, V., Denoel, P. & Poolman, J. T. 2011. Preclinical Evaluation Of The Pht Proteins As Potential Cross-Protective Pneumococcal Vaccine Antigens. *Infect Immun,* 79**,** 238-45.

Goldblatt, D., Hussain, M., Andrews, N., Ashton, L., Virta, C., Melegaro, A., Pebody, R., George, R., Soininen, A., Edmunds, J., Gay, N., Kayhty, H. & Miller, E. 2005. Antibody Responses To Nasopharyngeal Carriage Of Streptococcus Pneumoniae In Adults: A Longitudinal Household Study. *The Journal Of Infectious Diseases,* 192**,** 387-93.

Goldblatt, D., Ramakrishnan, M. & O'brien, K. 2013. Using The Impact Of Pneumococcal Vaccines On Nasopharyngeal Carriage To Aid Licensing And Vaccine Implementation; A Pneumocarr Meeting Report March 27-28, 2012, Geneva. *Vaccine,* 32**,** 146-52.

Goldenberg, H. B., Mccool, T. L. & Weiser, J. N. 2004. Cross-Reactivity Of Human Immunoglobulin G2 Recognizing Phosphorylcholine And Evidence For Protection Against Major Bacterial Pathogens Of The Human Respiratory Tract. *J Infect Dis,* 190**,** 1254-63.

Gong, H. L., Shi, Y., Zhou, L., Wu, C. P., Cao, P. Y., Tao, L., Xu, C., Hou, D. S. & Wang, Y. Z. 2013. The Composition Of Microbiome In Larynx And The Throat Biodiversity Between Laryngeal Squamous Cell Carcinoma Patients And Control Population. *Plos One,* 8**,** E66476.

Gor, D. O., Ding, X., Briles, D. E., Jacobs, M. R. & Greenspan, N. S. 2005. Relationship Between Surface Accessibility For Ppma, Psaa, And Pspa And

Antibody-Mediated Immunity To Systemic Infection By Streptococcus Pneumoniae. *Infect Immun,* 73**,** 1304-12.

Gordon, S. B., Chaponda, M., Walsh, A. L., Whitty, C. J., Gordon, M. A., Machili, C. E., Gilks, C. F., Boeree, M. J., Kampondeni, S., Read, R. C. & Molyneux, M. E. 2002. Pneumococcal Disease In Hiv-Infected Malawian Adults: Acute Mortality And Long-Term Survival. *Aids,* 16**,** 1409-17.

Gordon, S. B., Walsh, A. L., Chaponda, M., Gordon, M. A., Soko, D., Mbwvinji, M., Molyneux, M. E. & Read, R. C. 2000. Bacterial Meningitis In Malawian Adults: Pneumococcal Disease Is Common, Severe, And Seasonal. *Clin Infect Dis,* 31**,** 53-7.

Gotelli, N. & Colwell, R. 2001. Quantifying Biodiversity: Procedures And Pitfalls In Measurement And Comparison Of Species Richness. *Ecology Letters,* 4**,** 379-391.

Gould, J. M. & Weiser, J. N. 2001. Expression Of C-Reactive Protein In The Human Respiratory Tract. *Infection And Immunity,* 69**,** 1747-54.

Gould, J. M. & Weiser, J. N. 2002. The Inhibitory Effect Of C-Reactive Protein On Bacterial Phosphorylcholine Platelet-Activating Factor Receptor-Mediated Adherence Is Blocked By Surfactant. *J Infect Dis,* 186**,** 361-71.

Grada, A. & Weinbrecht, K. 2013. Next-Generation Sequencing: Methodology And Application. *The Journal Of Investigative Dermatology,* 133**,** E11.

Granat, S. M., Ollgren, J., Herva, E., Mia, Z., Auranen, K. & Makela, P. H. 2009. Epidemiological Evidence For Serotype-Independent Acquired Immunity To Pneumococcal Carriage. *J Infect Dis,* 200**,** 99-106.

Graninger, M., Nidetzky, B., Heinrichs, D. E., Whitfield, C. & Messner, P. 1999. Characterization Of Dtdp-4-Dehydrorhamnose 3,5-Epimerase And Dtdp-4-Dehydrorhamnose Reductase, Required For Dtdp-L-Rhamnose Biosynthesis In Salmonella Enterica Serovar Typhimurium Lt2. *J Biol Chem,* 274**,** 25069-77.

Gratten, M., Gratten, H., Poli, A., Carrad, E., Raymer, M. & Koki, G. 1986. Colonisation Of Haemophilus Influenzae And Streptococcus Pneumoniae In The Upper Respiratory Tract Of Neonates In Papua New Guinea: Primary Acquisition, Duration Of Carriage, And Relationship To Carriage In Mothers. *Biology Of The Neonate,* 50**,** 114-20.

Gratten, M., Manning, K., Dixon, J., Morey, F., Torzillo, P., Hanna, J., Erlich, J., Asche, V. & Riley, I. 1994. Upper Airway Carriage By Haemophilus Influenzae And Streptococcus Pneumoniae In Australian Aboriginal Children Hospitalised With Acute Lower Respiratory Infection. *Southeast Asian J Trop Med Public Health,* 25**,** 123-31.

Gray, B. M., Converse, G. M., 3rd & Dillon, H. C., Jr. 1980. Epidemiologic Studies Of Streptococcus Pneumoniae In Infants: Acquisition, Carriage, And Infection During The First 24 Months Of Life. *J Infect Dis,* 142**,** 923-33.

Greenberg, D., Givon-Lavi, N., Newman, N., Bar-Ziv, J. & Dagan, R. 2011. Nasopharyngeal Carriage Of Individual Streptococcus Pneumoniae Serotypes

During Pediatric Pneumonia As A Means To Estimate Serotype Disease Potential. *The Pediatric Infectious Disease Journal,* 30**,** 227-33.

Griffioen, A. W., Rijkers, G. T., Janssens-Korpela, P. & Zegers, B. J. 1991. Pneumococcal Polysaccharides Complexed With C3d Bind To Human B Lymphocytes Via Complement Receptor Type 2. *Infect Immun,* 59**,** 1839-45.

Griffioen, A. W., Toebes, E. A., Zegers, B. J. & Rijkers, G. T. 1992. Role Of Cr2 In The Human Adult And Neonatal In Vitro Antibody Response To Type 4 Pneumococcal Polysaccharide. *Cell Immunol,* 143**,** 11-22.

Griffiths, A. J. F., Miller, J. H., Suzuki, D. T. & Al., E. 2000. An Introduction To Genetic Analysis. New York W.H Freeman.

Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S. K., Sodergren, E., Methe, B., Desantis, T. Z., Petrosino, J. F., Knight, R. & Birren, B. W. 2011. Chimeric 16s Rrna Sequence Formation And Detection In Sanger And 454-Pyrosequenced Pcr Amplicons. *Genome Research,* 21**,** 494-504.

Hakenbeck, R., Kaminski, K., Konig, A., Van Der Linden, M., Paik, J., Reichmann, P. & Zahner, D. 1999. Penicillin-Binding Proteins In Beta-Lactam-Resistant Streptococcus Pneumoniae. *Microb Drug Resist,* 5**,** 91-9.

Hall-Stoodley, L., Costerton, J. W. & Stoodley, P. 2004. Bacterial Biofilms: From The Natural Environment To Infectious Diseases. *Nat Rev Microbiol,* 2**,** 95-108.

Hammerschmidt, S., Wolff, S., Hocke, A., Rosseau, S., Muller, E. & Rohde, M. 2005. Illustration Of Pneumococcal Polysaccharide Capsule During Adherence And Invasion Of Epithelial Cells. *Infect Immun,* 73**,** 4653-67.

Hammitt, L. L., Bruden, D. L., Butler, J. C., Baggett, H. C., Hurlburt, D. A., Reasonover, A. & Hennessy, T. W. 2006. Indirect Effect Of Conjugate Vaccine On Adult Carriage Of Streptococcus Pneumoniae: An Explanation Of Trends In Invasive Pneumococcal Disease. *J Infect Dis,* 193**,** 1487-94.

Hanage, W. P., Auranen, K., Syrjanen, R., Herva, E., Makela, P. H., Kilpi, T. & Spratt, B. G. 2004. Ability Of Pneumococcal Serotypes And Clones To Cause Acute Otitis Media: Implications For The Prevention Of Otitis Media By Conjugate Vaccines. *Infection And Immunity,* 72**,** 76-81.

Hanage, W. P., Fraser, C., Tang, J., Connor, T. R. & Corander, J. 2009. Hyper-Recombination, Diversity, And Antibiotic Resistance In Pneumococcus. *Science,* 324**,** 1454-7.

Hanage, W. P., Kaijalainen, T. H., Syrjanen, R. K., Auranen, K., Leinonen, M., Makela, P. H. & Spratt, B. G. 2005. Invasiveness Of Serotypes And Clones Of Streptococcus Pneumoniae Among Children In Finland. *Infection And Immunity,* 73**,** 431-5.

Hare, K. M., Morris, P., Smith-Vaughan, H. & Leach, A. J. 2008. Random Colony Selection Versus Colony Morphology For Detection Of Multiple Pneumococcal Serotypes In Nasopharyngeal Swabs. *Pediatr Infect Dis J,* 27**,** 178-80.

Harrison, L. M., Morris, J. A., Telford, D. R., Brown, S. M. & Jones, K. 1999. The Nasopharyngeal Bacterial Flora In Infancy: Effects Of Age, Gender, Season, Viral Upper Respiratory Tract Infection And Sleeping Position. *Fems Immunol Med Microbiol,* 25**,** 19-28.

Hatt, J. K. & Rather, P. N. 2008. Role Of Bacterial Biofilms In Urinary Tract Infections. *Curr Top Microbiol Immunol,* 322**,** 163-92.

Hausdorff, W. P., Bryant, J., Paradiso, P. R. & Siber, G. R. 2000. Which Pneumococcal Serogroups Cause The Most Invasive Disease: Implications For Conjugate Vaccine Formulation And Use, Part I. *Clinical Infectious Diseases : An Official Publication Of The Infectious Diseases Society Of America,* 30**,** 100-21.

Hauser, C., Aebi, S. & Muhlemann, K. 2004. An Internationally Spread Clone Of Streptococcus Pneumoniae Evolves From Low-Level To Higher-Level Penicillin Resistance By Uptake Of Penicillin-Binding Protein Gene Fragments From Nonencapsulated Pneumococci. *Antimicrobial Agents And Chemotherapy,* 48**,** 3563-6.

Hernani Mde, L., Ferreira, P. C., Ferreira, D. M., Miyaji, E. N., Ho, P. L. & Oliveira, M. L. 2011. Nasal Immunization Of Mice With Lactobacillus Casei Expressing The Pneumococcal Surface Protein C Primes The Immune System And Decreases Pneumococcal Nasopharyngeal Colonization In Mice. *Fems Immunol Med Microbiol,* 62**,** 263-72.

Hill, P. C., Akisanya, A., Sankareh, K., Cheung, Y. B., Saaka, M., Lahai, G., Greenwood, B. M. & Adegbola, R. A. 2006. Nasopharyngeal Carriage Of

Streptococcus Pneumoniae In Gambian Villagers. *Clinical Infectious Diseases : An Official Publication Of The Infectious Diseases Society Of America,* 43**,** 673-9.

Hill, P. C., Cheung, Y. B., Akisanya, A., Sankareh, K., Lahai, G., Greenwood, B. M. & Adegbola, R. A. 2008a. Nasopharyngeal Carriage Of Streptococcus Pneumoniae In Gambian Infants: A Longitudinal Study. *Clin Infect Dis,* 46**,** 807-14.

Hill, P. C., Cheung, Y. B., Akisanya, A., Sankareh, K., Lahai, G., Greenwood, B. M. & Adegbola, R. A. 2008b. Nasopharyngeal Carriage Of Streptococcus Pneumoniae In Gambian Infants: A Longitudinal Study. *Clinical Infectious Diseases : An Official Publication Of The Infectious Diseases Society Of America,* 46**,** 807-14.

Hiller, N. L., Ahmed, A., Powell, E., Martin, D. P., Eutsey, R., Earl, J., Janto, B., Boissy, R. J., Hogg, J., Barbadora, K., Sampath, R., Lonergan, S., Post, J. C., Hu, F. Z. & Ehrlich, G. D. 2010. Generation Of Genic Diversity Among Streptococcus Pneumoniae Strains Via Horizontal Gene Transfer During A Chronic Polyclonal Pediatric Infection. *Plos Pathog,* 6**,** E1001108.

Hilty, M., Qi, W., Brugger, S. D., Frei, L., Agyeman, P., Frey, P. M., Aebi, S. & Muhlemann, K. 2012. Nasopharyngeal Microbiota In Infants With Acute Otitis Media. *The Journal Of Infectious Diseases,* 205**,** 1048-55.

Hinds, J., Gould, K. A., Witney, A. A., Baldry, S. J., Lambertsen, L., Hannage, W. P., Antonio, M., Turner, P., Newton, R., Wernisch, L., Aanensen, D. M. & Bentley, S. 2009. Molecular Serotyping Of Streptococcus Pneumoniae: A

Microarray-Based Tool With Enhanced Utility For Isolate Typing, Novel Serotype Discovery, Non-Typeable Investigation, Multiple Carriage Detection, And Direct Analysis Of Nasopharyngeal Swabs [Abstract A-O5]. *9th European Meeting On The Molecular Biology Of The Pneumococcus (Europneumo). Bern, Switzerland.*

Hirschtick, R. E., Glassroth, J., Jordan, M. C., Wilcosky, T. C., Wallace, J. M., Kvale, P. A., Markowitz, N., Rosen, M. J., Mangura, B. T. & Hopewell, P. C. 1995. Bacterial Pneumonia In Persons Infected With The Human Immunodeficiency Virus. Pulmonary Complications Of Hiv Infection Study Group. *The New England Journal Of Medicine,* 333**,** 845-51.

Hirst, R. A., Kadioglu, A., O'callaghan, C. & Andrew, P. W. 2004. The Role Of Pneumolysin In Pneumococcal Pneumonia And Meningitis. *Clin Exp Immunol,* 138**,** 195-201.

Hoedemaekers, A., Schulin, T., Tonk, B., Melchers, W. J. & Sturm, P. D. 2006. Ventilator-Associated Pneumonia Caused By Dolosigranulum Pigrum. *J Clin Microbiol,* 44**,** 3461-2.

Hogberg, L., Geli, P., Ringberg, H., Melander, E., Lipsitch, M. & Ekdahl, K. 2007. Age- And Serogroup-Related Differences In Observed Durations Of Nasopharyngeal Carriage Of Penicillin-Resistant Pneumococci. *J Clin Microbiol,* 45**,** 948-52.

Holmes, A. R., Mcnab, R., Millsap, K. W., Rohde, M., Hammerschmidt, S., Mawdsley, J. L. & Jenkinson, H. F. 2001. The Pava Gene Of Streptococcus

Pneumoniae Encodes A Fibronectin-Binding Protein That Is Essential For Virulence. *Mol Microbiol,* 41**,** 1395-408.

Howden, R. 1976. Use Of Anaerobic Culture For The Improved Isolation Of Streptococcus Pneumoniae. *Journal Of Clinical Pathology,* 29**,** 50-3.

Hu, P., Brodie, E. L., Suzuki, Y., Mcadams, H. H. & Andersen, G. L. 2005. Whole-Genome Transcriptional Analysis Of Heavy Metal Stresses In Caulobacter Crescentus. *J Bacteriol,* 187**,** 8437-49.

Huang, S. S., Hinrichsen, V. L., Stevenson, A. E., Rifas-Shiman, S. L., Kleinman, K., Pelton, S. I., Lipsitch, M., Hanage, W. P., Lee, G. M. & Finkelstein, J. A. 2009. Continued Impact Of Pneumococcal Conjugate Vaccine On Carriage In Young Children. *Pediatrics,* 124**,** E1-11.

Huang, Y. J. & Lynch, S. V. 2011. The Emerging Relationship Between The Airway Microbiota And Chronic Respiratory Disease: Clinical Implications. *Expert Review Of Respiratory Medicine,* 5**,** 809-21.

Huang, Y. J., Nelson, C. E., Brodie, E. L., Desantis, T. Z., Baek, M. S., Liu, J., Woyke, T., Allgaier, M., Bristow, J., Wiener-Kronish, J. P., Sutherland, E. R., King, T. S., Icitovic, N., Martin, R. J., Calhoun, W. J., Castro, M., Denlinger, L. C., Dimango, E., Kraft, M., Peters, S. P., Wasserman, S. I., Wechsler, M. E., Boushey, H. A. & Lynch, S. V. 2011. Airway Microbiota And Bronchial Hyperresponsiveness In Patients With Suboptimally Controlled Asthma. *The Journal Of Allergy And Clinical Immunology,* 127**,** 372-381 E1-3.

Huebner, R. E., Dagan, R., Porath, N., Wasas, A. D. & Klugman, K. P. 2000a. Lack Of Utility Of Serotyping Multiple Colonies For Detection Of Simultaneous Nasopharyngeal Carriage Of Different Pneumococcal Serotypes. *Pediatr Infect Dis J,* 19**,** 1017-20.

Huebner, R. E., Wasas, A., Mushi, A., Mazhani, L. & Klugman, K. 1998. Nasopharyngeal Carriage And Antimicrobial Resistance In Isolates Of Streptococcus Pneumoniae And Haemophilus Influenzae Type B In Children Under 5 Years Of Age In Botswana. *Int J Infect Dis,* 3**,** 18-25.

Huebner, R. E., Wasas, A. D. & Klugman, K. P. 2000b. Prevalence Of Nasopharyngeal Antibiotic-Resistant Pneumococcal Carriage In Children Attending Private Paediatric Practices In Johannesburg. *S Afr Med J,* 90**,** 1116-21.

Hulgan, T., Shepherd, B. E., Raffanti, S. P., Fusco, J. S., Beckerman, R., Barkanic, G. & Sterling, T. R. 2007. Absolute Count And Percentage Of Cd4+ Lymphocytes Are Independent Predictors Of Disease Progression In Hiv-Infected Persons Initiating Highly Active Antiretroviral Therapy. *J Infect Dis,* 195**,** 425-31.

Huse, S. M., Welch, D. M., Morrison, H. G. & Sogin, M. L. 2010. Ironing Out The Wrinkles In The Rare Biosphere Through Improved Otu Clustering. *Environmental Microbiology,* 12**,** 1889-98.

Hussain, M., Melegaro, A., Pebody, R. G., George, R., Edmunds, W. J., Talukdar, R., Martin, S. A., Efstratiou, A. & Miller, E. 2005. A Longitudinal Household

Study Of Streptococcus Pneumoniae Nasopharyngeal Carriage In A Uk Setting. *Epidemiology And Infection,* 133**,** 891-8.

Huyghe, A., Francois, P. & Schrenzel, J. 2009. Characterization Of Microbial Pathogens By Dna Microarrays. *Infection, Genetics And Evolution : Journal Of Molecular Epidemiology And Evolutionary Genetics In Infectious Diseases,* 9**,** 987-95.

Hyams, C., Camberlein, E., Cohen, J. M., Bax, K. & Brown, J. S. 2010. The Streptococcus Pneumoniae Capsule Inhibits Complement Activity And Neutrophil Phagocytosis By Multiple Mechanisms. *Infection And Immunity,* 78**,** 704-15.

Iannelli, F., Chiavolini, D., Ricci, S., Oggioni, M. R. & Pozzi, G. 2004. Pneumococcal Surface Protein C Contributes To Sepsis Caused By Streptococcus Pneumoniae In Mice. *Infect Immun,* 72**,** 3077-80.

Ingram, D. L., Collier, A. M., Pendergrass, E. & King, S. H. 1979. Methods For Serotyping Nasopharyngeal Isolates Of Haemophilus Influenzae: Slide Agglutination, Quellung Reaction, Countercurrent Immunoelectrophoresis, Latex Agglutination, And Antiserum Agar. *J Clin Microbiol,* 9**,** 570-4.

Isaacman, D. J., Fletcher, M. A., Fritzell, B., Ciuryla, V. & Schranz, J. 2007. Indirect Effects Associated With Widespread Vaccination Of Infants With Heptavalent Pneumococcal Conjugate Vaccine (Pcv7; Prevnar). *Vaccine,* 25**,** 2420-7.

Isaacman, D. J., Mcintosh, E. D. & Reinert, R. R. 2010. Burden Of Invasive Pneumococcal Disease And Serotype Distribution Among Streptococcus

Pneumoniae Isolates In Young Children In Europe: Impact Of The 7-Valent Pneumococcal Conjugate Vaccine And Considerations For Future Conjugate Vaccines. *Int J Infect Dis,* 14**,** E197-209.

Ito, A., Taniuchi, A., May, T., Kawata, K. & Okabe, S. 2009. Increased Antibiotic Resistance Of Escherichia Coli In Mature Biofilms. *Appl Environ Microbiol,* 75**,** 4093-100.

Jacoby, P., Watson, K., Bowman, J., Taylor, A., Riley, T. V., Smith, D. W. & Lehmann, D. 2007. Modelling The Co-Occurrence Of Streptococcus Pneumoniae With Other Bacterial And Viral Pathogens In The Upper Respiratory Tract. *Vaccine,* 25**,** 2458-64.

Jambo, K. C., Sepako, E., Heyderman, R. S. & Gordon, S. B. 2010. Potential Role For Mucosally Active Vaccines Against Pneumococcal Pneumonia. *Trends Microbiol,* 18**,** 81-9.

Janda, J. M. & Abbott, S. L. 2007. 16s Rrna Gene Sequencing For Bacterial Identification In The Diagnostic Laboratory: Pluses, Perils, And Pitfalls. *J Clin Microbiol,* 45**,** 2761-4.

Jedrzejas, M. J. 2001. Pneumococcal Virulence Factors: Structure And Function. *Microbiol Mol Biol Rev,* 65**,** 187-207 ; First Page, Table Of Contents.

Johnson, H. L., Deloria-Knoll, M., Levine, O. S., Stoszek, S. K., Freimanis Hance, L., Reithinger, R., Muenz, L. R. & O'brien, K. L. 2010. Systematic Evaluation Of Serotypes Causing Invasive Pneumococcal Disease Among Children Under Five: The Pneumococcal Global Serotype Project. *Plos Med,* 7**,** 1-13.

Johnston, J. W., Briles, D. E., Myers, L. E. & Hollingshead, S. K. 2006. Mn2+-Dependent Regulation Of Multiple Genes In Streptococcus Pneumoniae Through Psar And The Resultant Impact On Virulence. *Infect Immun,* 74**,** 1171-80.

Joloba, M. L., Bajaksouzian, S., Palavecino, E., Whalen, C. & Jacobs, M. R. 2001. High Prevalence Of Carriage Of Antibiotic-Resistant Streptococcus Pneumoniae In Children In Kampala Uganda. *Int J Antimicrob Agents,* 17**,** 395-400.

Jones, J. B., Bouzar, H., Stall, R. E., Almira, E. C., Roberts, P. D., Bowen, B. W., Sudberry, J., Strickler, P. M. & Chun, J. 2000. Systematic Analysis Of Xanthomonads (Xanthomonas Spp.) Associated With Pepper And Tomato Lesions. *Int J Syst Evol Microbiol,* 50 Pt 3**,** 1211-9.

Jordano, Q., Falco, V., Almirante, B., Planes, A. M., Del Valle, O., Ribera, E., Len, O., Pigrau, C. & Pahissa, A. 2004. Invasive Pneumococcal Disease In Patients Infected With Hiv: Still A Threat In The Era Of Highly Active Antiretroviral Therapy. *Clinical Infectious Diseases : An Official Publication Of The Infectious Diseases Society Of America,* 38**,** 1623-8.

Joyce, E. A., Popper, S. J. & Falkow, S. 2009. Streptococcus Pneumoniae Nasopharyngeal Colonization Induces Type I Interferons And Interferon-Induced Gene Expression. *Bmc Genomics,* 10**,** 404.

Kadioglu, A., Gingles, N. A., Grattan, K., Kerr, A., Mitchell, T. J. & Andrew, P. W. 2000. Host Cellular Immune Response To Pneumococcal Lung Infection In Mice. *Infection And Immunity,* 68**,** 492-501.

Kadioglu, A., Weiser, J. N., Paton, J. C. & Andrew, P. W. 2008. The Role Of Streptococcus Pneumoniae Virulence Factors In Host Respiratory Colonization And Disease. *Nat Rev Microbiol,* 6**,** 288-301.

Kaltoft, M. S., Skov Sorensen, U. B., Slotved, H. C. & Konradsen, H. B. 2008. An Easy Method For Detection Of Nasopharyngeal Carriage Of Multiple Streptococcus Pneumoniae Serotypes. *J Microbiol Methods,* 75**,** 540-4.

Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S., Watanabe, A., Idesawa, K., Ishikawa, A., Kawashima, K., Kimura, T., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpo, S., Sugimoto, M., Takeuchi, C., Yamada, M. & Tabata, S. 2000. Complete Genome Structure Of The Nitrogen-Fixing Symbiotic Bacterium Mesorhizobium Loti. *Dna Res,* 7**,** 331-8.

Karlowsky, J. A., Thornsberry, C., Jones, M. E., Evangelista, A. T., Critchley, I. A. & Sahm, D. F. 2003. Factors Associated With Relative Rates Of Antimicrobial Resistance Among Streptococcus Pneumoniae In The United States: Results From The Trust Surveillance Program (1998-2002). *Clin Infect Dis,* 36**,** 963-70.

Karstaedt, A. S., Khoosal, M. & Crewe-Brown, H. H. 2000. Pneumococcal Bacteremia During A Decade In Children In Soweto, South Africa. *Pediatr Infect Dis J,* 19**,** 454-7.

Karstaedt, A. S., Khoosal, M. & Crewe-Brown, H. H. 2001. Pneumococcal Bacteremia In Adults In Soweto, South Africa, During The Course Of A Decade. *Clin Infect Dis,* 33**,** 610-4.

Kellogg, J. A., Bankert, D. A., Elder, C. J., Gibbs, J. L. & Smith, M. C. 2001. Identification Of Streptococcus Pneumoniae Revisited. *J Clin Microbiol,* 39**,** 3373-5.

Khan, M. N. & Pichichero, M. E. 2012. Vaccine Candidates Phtd And Phte Of Streptococcus Pneumoniae Are Adhesins That Elicit Functional Antibodies In Humans. *Vaccine,* 30**,** 2900-7.

Kim, Y. J., Nguyen, N. L., Weon, H. Y. & Yang, D. C. 2013. Sediminibacterium Ginsengisoli Sp. Nov., Isolated From Soil Of A Ginseng Field, And Emended Descriptions Of The Genus Sediminibacterium And Of Sediminibacterium Salmoneum. *Int J Syst Evol Microbiol,* 63**,** 905-12.

King, S. J., Hippe, K. R., Gould, J. M., Bae, D., Peterson, S., Cline, R. T., Fasching, C., Janoff, E. N. & Weiser, J. N. 2004. Phase Variable Desialylation Of Host Proteins That Bind To Streptococcus Pneumoniae In Vivo And Protect The Airway. *Molecular Microbiology,* 54**,** 159-71.

King, S. J., Hippe, K. R. & Weiser, J. N. 2006a. Deglycosylation Of Human Glycoconjugates By The Sequential Activities Of Exoglycosidases Expressed By Streptococcus Pneumoniae. *Molecular Microbiology,* 59**,** 961-74.

King, S. J., Hippe, K. R. & Weiser, J. N. 2006b. Deglycosylation Of Human Glycoconjugates By The Sequential Activities Of Exoglycosidases Expressed By Streptococcus Pneumoniae. *Mol Microbiol,* 59**,** 961-74.

Kluytmans, J., Van Belkum, A. & Verbrugh, H. 1997. Nasal Carriage Of Staphylococcus Aureus: Epidemiology, Underlying Mechanisms, And Associated Risks. *Clin Microbiol Rev,* 10**,** 505-20.

Ko, K. S., Baek, J. Y. & Song, J. H. 2013. Capsular Gene Sequences And Genotypes Of "Serotype 6e" Streptococcus Pneumoniae Isolates. *J Clin Microbiol,* 51**,** 3395-9.

Kong, H. H., Oh, J., Deming, C., Conlan, S., Grice, E. A., Beatson, M. A., Nomicos, E., Polley, E. C., Komarow, H. D., Murray, P. R., Turner, M. L. & Segre, J. A. 2012. Temporal Shifts In The Skin Microbiome Associated With Disease Flares And Treatment In Children With Atopic Dermatitis. *Genome Research,* 22**,** 850-9.

Konradsen, H. B. 2005. Validation Of Serotyping Of Streptococcus Pneumoniae In Europe. *Vaccine,* 23**,** 1368-73.

Kukavica-Ibrulj, I., Hamelin, M. E., Prince, G. A., Gagnon, C., Bergeron, Y., Bergeron, M. G. & Boivin, G. 2009. Infection With Human Metapneumovirus Predisposes Mice To Severe Pneumococcal Pneumonia. *Journal Of Virology,* 83**,** 1341-9.

Kwambana, B. A., Barer, M. R., Bottomley, C., Adegbola, R. A. & Antonio, M. 2011. Early Acquisition And High Nasopharyngeal Co-Colonisation By Streptococcus Pneumoniae And Three Respiratory Pathogens Amongst Gambian New-Borns And Infants. *Bmc Infectious Diseases,* 11**,** 175.

Kyaw, M. H., Christie, P., Jones, I. G. & Campbell, H. 2002. The Changing Epidemiology Of Bacterial Meningitis And Invasive Non-Meningitic

Bacterial Disease In Scotland During The Period 1983-99. *Scand J Infect Dis,* 34**,** 289-98.

Laaberki, M. H., Pfeffer, J., Clarke, A. J. & Dworkin, J. 2011. O-Acetylation Of Peptidoglycan Is Required For Proper Cell Separation And S-Layer Anchoring In Bacillus Anthracis. *The Journal Of Biological Chemistry,* 286**,** 5278-88.

Laible, G., Spratt, B. G. & Hakenbeck, R. 1991. Interspecies Recombinational Events During The Evolution Of Altered Pbp 2x Genes In Penicillin-Resistant Clinical Isolates Of Streptococcus Pneumoniae. *Mol Microbiol,* 5**,** 1993-2002.

Lau, C. & Muula, A. S. 2004. Hiv/Aids In Sub-Saharan Africa. *Croatian Medical Journal,* 45**,** 402-14.

Laukeland, H., Bergh, K. & Bevanger, L. 2002. Posttrabeculectomy Endophthalmitis Caused By Moraxella Nonliquefaciens. *J Clin Microbiol,* 40**,** 2668-770.

Le, T., Wright, E. J., Smith, D. M., He, W., Catano, G., Okulicz, J. F., Young, J. A., Clark, R. A., Richman, D. D., Little, S. J. & Ahuja, S. K. 2013. Enhanced Cd4+ T-Cell Recovery With Earlier Hiv-1 Antiretroviral Therapy. *The New England Journal Of Medicine,* 368**,** 218-30.

Lebien, T. W. & Tedder, T. F. 2008. B Lymphocytes: How They Develop And Function. *Blood,* 112**,** 1570-80.

Lecuyer, H., Audibert, J., Bobigny, A., Eckert, C., Janniere-Nartey, C., Buu-Hoi, A., Mainardi, J. L. & Podglajen, I. 2007. Dolosigranulum Pigrum Causing Nosocomial Pneumonia And Septicemia. *J Clin Microbiol,* 45**,** 3474-5.

Lee, G. M., Kleinman, K., Pelton, S. I., Hanage, W. P., Huang, S. S., Lakoma, M., Dutta-Linn, M., Croucher, N. J., Stevenson, A. & Finkelstein, J. A. 2013a. Impact Of 13-Valent Pneumococcal Conjugate Vaccination On Streptococcus Pneumoniae Carriage In Young Children In Massachusetts. *Journal Of The Pediatric Infectious Diseases Society***,** 1-10.

Lee, J., Lee, T. K., Woo, S. G., Baek, G. S. & Park, J. 2013b. In-Depth Characterization Of Wastewater Bacterial Community In Response To Algal Growth Using Pyrosequencing. *Journal Of Microbiology And Biotechnology*.

Lee, T. M., Sodhi, N. S. & Prawiradilaga, D. M. 2007. The Importance Of Protected Areas For The Forest And Endemic Avifauna Of Sulawesi (Indonesia). *Ecol Appl,* 17**,** 1727-41.

Lemon, K. P., Klepac-Ceraj, V., Schiffer, H. K., Brodie, E. L., Lynch, S. V. & Kolter, R. 2010. Comparative Analyses Of The Bacterial Microbiota Of The Human Nostril And Oropharynx. *Mbio,* 1**,** 1-9.

Leung, M. H., Bryson, K., Freystatter, K., Pichon, B., Edwards, G., Charalambous, B. M. & Gillespie, S. H. 2012. Sequetyping: Serotyping Streptococcus Pneumoniae By A Single Pcr Sequencing Strategy. *Journal Of Clinical Microbiology,* 50**,** 2419-27.

Levy, R. & Borenstein, E. 2013. Metabolic Modeling Of Species Interaction In The Human Microbiome Elucidates Community-Level Assembly Rules. *Proc Natl Acad Sci U S A,* 110**,** 12804-9.

Li, H. & Durbin, R. 2009. Fast And Accurate Short Read Alignment With Burrows-Wheeler Transform. *Bioinformatics,* 25**,** 1754-60.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. 2009. The Sequence Alignment/Map Format And Samtools. *Bioinformatics,* 25**,** 2078-9.

Li, Q., Hobbs, M. & Reeves, P. R. 2003. The Variation Of Dtdp-L-Rhamnose Pathway Genes In Vibrio Cholerae. *Microbiology,* 149**,** 2463-74.

Ling, Z., Liu, X., Luo, Y., Yuan, L., Nelson, K. E., Wang, Y., Xiang, C. & Li, L. 2013. Pyrosequencing Analysis Of The Human Microbiota Of Healthy Chinese Undergraduates. *Bmc Genomics,* 14**,** 390.

Link-Gelles, R., Taylor, T. & Moore, M. R. 2013. Forecasting Invasive Pneumococcal Disease Trends After The Introduction Of 13-Valent Pneumococcal Conjugate Vaccine In The United States, 2010-2020. *Vaccine,* 31**,** 2572-7.

Lipsitch, M. 1999. Bacterial Vaccines And Serotype Replacement: Lessons From Haemophilus Influenzae And Prospects For Streptococcus Pneumoniae. *Emerg Infect Dis,* 5**,** 336-45.

Lloyd-Evans, N., O'dempsey, T. J., Baldeh, I., Secka, O., Demba, E., Todd, J. E., Mcardle, T. F., Banya, W. S. & Greenwood, B. M. 1996. Nasopharyngeal

Carriage Of Pneumococci In Gambian Children And In Their Families. *The Pediatric Infectious Disease Journal,* 15**,** 866-71.

Llull, D., Munoz, R., Lopez, R. & Garcia, E. 1999. A Single Gene (Tts) Located Outside The Cap Locus Directs The Formation Of Streptococcus Pneumoniae Type 37 Capsular Polysaccharide. Type 37 Pneumococci Are Natural, Genetically Binary Strains. *J Exp Med,* 190**,** 241-51.

Long, J. P., Tong, H. H. & Demaria, T. F. 2004. Immunization With Native Or Recombinant Streptococcus Pneumoniae Neuraminidase Affords Protection In The Chinchilla Otitis Media Model. *Infect Immun,* 72**,** 4309-13.

Lopez, D., Vlamakis, H. & Kolter, R. 2010. Biofilms. *Cold Spring Harb Perspect Biol,* 2**,** A000398.

Lottenbach, K. R., Mink, C. M., Barenkamp, S. J., Anderson, E. L., Homan, S. M. & Powers, D. C. 1999. Age-Associated Differences In Immunoglobulin G1 (Igg1) And Igg2 Subclass Antibodies To Pneumococcal Polysaccharides Following Vaccination. *Infect Immun,* 67**,** 4935-8.

Loughlin, A. M., Hsu, K., Silverio, A. L., Marchant, C. D. & Pelton, S. I. 2014. Direct And Indirect Effects Of Pcv13 On Nasopharyngeal Carriage Of Pcv13 Unique Pneumococcal Serotypes In Massachusetts' Children. *Pediatr Infect Dis J,* 33**,** 504-10.

Lu, Y. J., Gross, J., Bogaert, D., Finn, A., Bagrade, L., Zhang, Q., Kolls, J. K., Srivastava, A., Lundgren, A., Forte, S., Thompson, C. M., Harney, K. F., Anderson, P. W., Lipsitch, M. & Malley, R. 2008. Interleukin-17a Mediates

Acquired Immunity To Pneumococcal Colonization. *Plos Pathog,* 4**,** E1000159.

Lu, Y. J., Leite, L., Goncalves, V. M., Dias Wde, O., Liberman, C., Fratelli, F., Alderson, M., Tate, A., Maisonneuve, J. F., Robertson, G., Graca, R., Sayeed, S., Thompson, C. M., Anderson, P. & Malley, R. 2010. Gmp-Grade Pneumococcal Whole-Cell Vaccine Injected Subcutaneously Protects Mice From Nasopharyngeal Colonization And Fatal Aspiration-Sepsis. *Vaccine,* 28**,** 7468-75.

Lysenko, E., Richards, J. C., Cox, A. D., Stewart, A., Martin, A., Kapoor, M. & Weiser, J. N. 2000. The Position Of Phosphorylcholine On The Lipopolysaccharide Of Haemophilus Influenzae Affects Binding And Sensitivity To C-Reactive Protein-Mediated Killing. *Mol Microbiol,* 35**,** 234-45.

Lysenko, E. S., Ratner, A. J., Nelson, A. L. & Weiser, J. N. 2005. The Role Of Innate Immune Responses In The Outcome Of Interspecies Competition For Colonization Of Mucosal Surfaces. *Plos Pathogens,* 1**,** E1.

Mackenzie, G. A., Carapetis, J. R., Leach, A. J. & Morris, P. S. 2009. Pneumococcal Vaccination And Otitis Media In Australian Aboriginal Infants: Comparison Of Two Birth Cohorts Before And After Introduction Of Vaccination. *Bmc Pediatr,* 9**,** 14.

Mackenzie, G. A., Leach, A. J., Carapetis, J. R., Fisher, J. & Morris, P. S. 2010. Epidemiology Of Nasopharyngeal Carriage Of Respiratory Bacterial

Pathogens In Children And Adults: Cross-Sectional Surveys In A Population With High Rates Of Pneumococcal Disease. *Bmc Infect Dis,* 10**,** 304.

Madhi, S. A., Adrian, P., Kuwanda, L., Cutland, C., Albrich, W. C. & Klugman, K. P. 2007a. Long-Term Effect Of Pneumococcal Conjugate Vaccine On Nasopharyngeal Colonization By Streptococcus Pneumoniae--And Associated Interactions With Staphylococcus Aureus And Haemophilus Influenzae Colonization--In Hiv-Infected And Hiv-Uninfected Children. *J Infect Dis,* 196**,** 1662-6.

Madhi, S. A., Adrian, P., Kuwanda, L., Cutland, C., Albrich, W. C. & Klugman, K. P. 2007b. Long-Term Effect Of Pneumococcal Conjugate Vaccine On Nasopharyngeal Colonization By Streptococcus Pneumoniae--And Associated Interactions With Staphylococcus Aureus And Haemophilus Influenzae Colonization--In Hiv-Infected And Hiv-Uninfected Children. *The Journal Of Infectious Diseases,* 196**,** 1662-6.

Mah, T. F. & O'toole, G. A. 2001. Mechanisms Of Biofilm Resistance To Antimicrobial Agents. *Trends Microbiol,* 9**,** 34-9.

Maiden, M. C., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D. A., Feavers, I. M., Achtman, M. & Spratt, B. G. 1998. Multilocus Sequence Typing: A Portable Approach To The Identification Of Clones Within Populations Of Pathogenic Microorganisms. *Proceedings Of The National Academy Of Sciences Of The United States Of America,* 95**,** 3140-5.

Malley, R. & Anderson, P. W. 2012. Serotype-Independent Pneumococcal Experimental Vaccines That Induce Cellular As Well As Humoral Immunity. *Proc Natl Acad Sci U S A,* 109**,** 3623-7.

Malley, R., Henneke, P., Morse, S. C., Cieslewicz, M. J., Lipsitch, M., Thompson, C. M., Kurt-Jones, E., Paton, J. C., Wessels, M. R. & Golenbock, D. T. 2003. Recognition Of Pneumolysin By Toll-Like Receptor 4 Confers Resistance To Pneumococcal Infection. *Proc Natl Acad Sci U S A,* 100**,** 1966-71.

Malley, R., Trzcinski, K., Srivastava, A., Thompson, C. M., Anderson, P. W. & Lipsitch, M. 2005. Cd4+ T Cells Mediate Antibody-Independent Acquired Immunity To Pneumococcal Colonization. *Proc Natl Acad Sci U S A,* 102**,** 4848-53.

Manco, S., Hernon, F., Yesilkaya, H., Paton, J. C., Andrew, P. W. & Kadioglu, A. 2006. Pneumococcal Neuraminidases A And B Both Have Essential Roles During Infection Of The Respiratory Tract And Sepsis. *Infect Immun,* 74**,** 4014-20.

Mao, D. P., Zhou, Q., Chen, C. Y. & Quan, Z. X. 2012. Coverage Evaluation Of Universal Bacterial Primers Using The Metagenomic Datasets. *Bmc Microbiology,* 12**,** 66.

Marchesi, J. R., Dutilh, B. E., Hall, N., Peters, W. H., Roelofs, R., Boleij, A. & Tjalsma, H. 2011. Towards The Human Colorectal Cancer Microbiome. *Plos One,* 6**,** E20447.

Marchisio, P., Claut, L., Rognoni, A., Esposito, S., Passali, D., Bellussi, L., Drago, L., Pozzi, G., Mannelli, S., Schito, G. & Principi, N. 2003. Differences In

Nasopharyngeal Bacterial Flora In Children With Nonsevere Recurrent Acute Otitis Media And Chronic Otitis Media With Effusion: Implications For Management. *Pediatr Infect Dis J,* 22**,** 262-8.

Mardis, E. R. 2008. Next-Generation Dna Sequencing Methods. *Annual Review Of Genomics And Human Genetics,* 9**,** 387-402.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L., Jarvie, T. P., Jirage, K. B., Kim, J. B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., Mcdade, K. E., Mckenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P., Begley, R. F. & Rothberg, J. M. 2005. Genome Sequencing In Microfabricated High-Density Picolitre Reactors. *Nature,* 437**,** 376-80.

Marimon, J. M., Perez-Trallero, E., Ercibengoa, M., Gonzalez, A. & Fenoll, A. 2006. Molecular Epidemiology And Variants Of The Multidrug-Resistant Streptococcus Pneumoniae Spain14-5 International Clone Among Spanish Clinical Isolates. *The Journal Of Antimicrobial Chemotherapy,* 57**,** 654-60.

Marks, L. R., Reddinger, R. M. & Hakansson, A. P. 2012. High Levels Of Genetic Recombination During Nasopharyngeal Carriage And Biofilm Formation In Streptococcus Pneumoniae. *Mbio,* 3**,** 1-13.

Marsh, R., Smith-Vaughan, H., Hare, K. M., Binks, M., Kong, F., Warning, J., Gilbert, G. L., Morris, P. & Leach, A. J. 2010. The Nonserotypeable Pneumococcus: Phenotypic Dynamics In The Era Of Anticapsular Vaccines. *Journal Of Clinical Microbiology,* 48**,** 831-5.

Marsh, R. L., Smith-Vaughan, H., Beissbarth, J., Hare, K., Kennedy, M., Wigger, C., Mellon, G., Stubbs, E., Gadil, J. R., Pettit, A., Mackenzie, G., Tipakalippa, P., Morris, P. S. & Leach, A. J. 2007. Molecular Characterisation Of Pneumococcal Serotype 16f: Established Predominant Carriage And Otitis Media Serotype In The 7vpcv Era. *Vaccine,* 25**,** 2434-6.

Martner, A., Skovbjerg, S., Paton, J. C. & Wold, A. E. 2009. Streptococcus Pneumoniae Autolysis Prevents Phagocytosis And Production Of Phagocyte-Activating Cytokines. *Infect Immun,* 77**,** 3826-37.

Marton, A., Gulyas, M., Munoz, R. & Tomasz, A. 1991. Extremely High Incidence Of Antibiotic Resistance In Clinical Isolates Of Streptococcus Pneumoniae In Hungary. *J Infect Dis,* 163**,** 542-8.

Mavroidi, A., Aanensen, D. M., Godoy, D., Skovsted, I. C., Kaltoft, M. S., Reeves, P. R., Bentley, S. D. & Spratt, B. G. 2007. Genetic Relatedness Of The Streptococcus Pneumoniae Capsular Biosynthetic Loci. *J Bacteriol,* 189**,** 7841-55.

Mbelle, N., Huebner, R. E., Wasas, A. D., Kimura, A., Chang, I. & Klugman, K. P. 1999. Immunogenicity And Impact On Nasopharyngeal Carriage Of A Nonavalent Pneumococcal Conjugate Vaccine. *J Infect Dis,* 180**,** 1171-6.

Mcallister, L. J., Tseng, H. J., Ogunniyi, A. D., Jennings, M. P., Mcewan, A. G. & Paton, J. C. 2004. Molecular Analysis Of The Psa Permease Complex Of Streptococcus Pneumoniae. *Mol Microbiol,* 53**,** 889-901.

Mccool, T. L., Cate, T. R., Moy, G. & Weiser, J. N. 2002. The Immune Response To Pneumococcal Proteins During Experimental Human Carriage. *J Exp Med,* 195**,** 359-65.

Mccool, T. L. & Weiser, J. N. 2004. Limited Role Of Antibody In Clearance Of Streptococcus Pneumoniae In A Murine Model Of Colonization. *Infect Immun,* 72**,** 5807-13.

Mccullers, J. A. 2006. Insights Into The Interaction Between Influenza Virus And Pneumococcus. *Clinical Microbiology Reviews,* 19**,** 571-82.

Mccullers, J. A. & Bartmess, K. C. 2003. Role Of Neuraminidase In Lethal Synergism Between Influenza Virus And Streptococcus Pneumoniae. *The Journal Of Infectious Diseases,* 187**,** 1000-9.

Mcdaniel, L. S., Sheffield, J. S., Delucchi, P. & Briles, D. E. 1991. Pspa, A Surface Protein Of Streptococcus Pneumoniae, Is Capable Of Eliciting Protection Against Pneumococci Of More Than One Capsular Type. *Infect Immun,* 59**,** 222-8.

Mcellistrem, M. C. 2009. Genetic Diversity Of The Pneumococcal Capsule: Implications For Molecular-Based Serotyping. *Future Microbiol,* 4**,** 857-65.

Mchardy, I. H., Xli, X., Tong, M., Ruegger, P., Jacobs, J., Borneman, J., Anton, P. & Braun, J. 2013. Hiv Infection Is Associated With Compositional And Functional Shifts In The Rectal Mucosal Microbiota. *Microbiome,* 1**,** 12.

Mcnamee, L. A. & Harmsen, A. G. 2006. Both Influenza-Induced Neutrophil Dysfunction And Neutrophil-Independent Mechanisms Contribute To Increased Susceptibility To A Secondary Streptococcus Pneumoniae Infection. *Infection And Immunity,* 74**,** 6707-21.

Metzker, M. L. 2010. Sequencing Technologies - The Next Generation. *Nature Reviews. Genetics,* 11**,** 31-46.

Miller, E., Andrews, N. J., Waight, P. A., Slack, M. P. & George, R. C. 2011. Herd Immunity And Serotype Replacement 4 Years After Seven-Valent Pneumococcal Conjugate Vaccination In England And Wales: An Observational Cohort Study. *Lancet Infect Dis,* 11**,** 760-8.

Mirza, S., Wilson, L., Benjamin, W. H., Jr., Novak, J., Barnes, S., Hollingshead, S. K. & Briles, D. E. 2011. Serine Protease Prta From Streptococcus Pneumoniae Plays A Role In The Killing Of S. Pneumoniae By Apolactoferrin. *Infect Immun,* 79**,** 2440-50.

Miyaji, E. N., Dias, W. O., Gamberini, M., Gebara, V. C., Schenkman, R. P., Wild, J., Riedl, P., Reimann, J., Schirmbeck, R. & Leite, L. C. 2001. Psaa (Pneumococcal Surface Adhesin A) And Pspa (Pneumococcal Surface Protein A) Dna Vaccines Induce Humoral And Cellular Immune Responses Against Streptococcus Pneumoniae. *Vaccine,* 20**,** 805-12.

Mlst[Database] Available At: Http://Spneumoniae.Mlst.Net.

Mook-Kanamori, B. B., Geldhoff, M., Van Der Poll, T. & Van De Beek, D. 2011. Pathogenesis And Pathophysiology Of Pneumococcal Meningitis. *Clinical Microbiology Reviews,* 24**,** 557-91.

Moore, D. M., Hogg, R. S., Yip, B., Craib, K., Wood, E. & Montaner, J. S. 2006. Cd4 Percentage Is An Independent Predictor Of Survival In Patients Starting Antiretroviral Therapy With Absolute Cd4 Cell Counts Between 200 And 350 Cells/Microl. *Hiv Med,* 7**,** 383-8.

Morais, L., Carvalho Mda, G., Roca, A., Flannery, B., Mandomando, I., Soriano-Gabarro, M., Sigauque, B., Alonso, P. & Beall, B. 2007. Sequential Multiplex Pcr For Identifying Pneumococcal Capsular Serotypes From South-Saharan African Clinical Isolates. *J Med Microbiol,* 56**,** 1181-4.

Morgan, X. C. & Huttenhower, C. 2012. Chapter 12: Human Microbiome Analysis. *Plos Computational Biology,* 8**,** E1002808.

Morona, J. K., Morona, R. & Paton, J. C. 1999. Comparative Genetics Of Capsular Polysaccharide Biosynthesis In Streptococcus Pneumoniae Types Belonging To Serogroup 19. *Journal Of Bacteriology,* 181**,** 5355-64.

Mortensen, R. F. & Duszkiewicz, J. A. 1977. Mediation Of Crp-Dependent Phagocytosis Through Mouse Macrophage Fc-Receptors. *Journal Of Immunology,* 119**,** 1611-6.

Morton, D. B. & Griffiths, P. H. 1985. Guidelines On The Recognition Of Pain, Distress And Discomfort In Experimental Animals And An Hypothesis For Assessment. *The Veterinary Record,* 116**,** 431-6.

Mukerji, R., Mirza, S., Roche, A. M., Widener, R. W., Croney, C. M., Rhee, D. K., Weiser, J. N., Szalai, A. J. & Briles, D. E. 2012. Pneumococcal Surface Protein A Inhibits Complement Deposition On The Pneumococcal Surface By Competing With The Binding Of C-Reactive Protein To Cell-Surface Phosphocholine. *J Immunol,* 189**,** 5327-35.

Mulholland, K. & Satzke, C. 2012. Serotype Replacement After Pneumococcal Vaccination. *Lancet,* 379**,** 1387; Author Reply 1388-9.

Munoz-Price, L. S. & Weinstein, R. A. 2008. Acinetobacter Infection. *N Engl J Med,* 358**,** 1271-81.

Murray, P. R. 1979. Modification Of The Bile Solubility Test For Rapid Identification Of Streptococcus Pneumoniae. *J Clin Microbiol,* 9**,** 290-1.

Nagel, M. & Andreesen, J. R. 1991. Bacillus Niacini Sp. Nov. A Nicotinate-Metabolizing Mesophile Isolated From Soil. *International Journal Of Systematic Bacteriology,* 41**,** 134-139.

Naseeb, S. & Delneri, D. 2012. Impact Of Chromosomal Inversions On The Yeast Dal Cluster. *Plos One,* 7**,** E42022.

Nazli, A., Chan, O., Dobson-Belaire, W. N., Ouellet, M., Tremblay, M. J., Gray-Owen, S. D., Arsenault, A. L. & Kaushic, C. 2010. Exposure To Hiv-1 Directly Impairs Mucosal Epithelial Barrier Integrity Allowing Microbial Translocation. *Plos Pathog,* 6**,** E1000852.

Nei, M. 1972. Genetic Distance Between Populations. *The American Naturalist,* 106**,** 283-292.

Nelson, A. L., Roche, A. M., Gould, J. M., Chim, K., Ratner, A. J. & Weiser, J. N. 2007a. Capsule Enhances Pneumococcal Colonization By Limiting Mucus-Mediated Clearance. *Infection And Immunity,* 75**,** 83-90.

Nelson, A. L., Roche, A. M., Gould, J. M., Chim, K., Ratner, A. J. & Weiser, J. N. 2007b. Capsule Enhances Pneumococcal Colonization By Limiting Mucus-Mediated Clearance. *Infect Immun,* 75**,** 83-90.

Newton, R., Hinds, J. & Wernisch, L. 2011. Empirical Bayesian Models For Analysing Molecular Serotyping Microarrays. *Bmc Bioinformatics,* 12**,** 88.

Nicoletti, C., Brandileone, M. C., Guerra, M. L. & Levin, A. S. 2007. Prevalence, Serotypes, And Risk Factors For Pneumococcal Carriage Among Hiv-Infected Adults. *Diagn Microbiol Infect Dis,* 57**,** 259-65.

Nikaido, H. 2009. Multidrug Resistance In Bacteria. *Annu Rev Biochem,* 78**,** 119-46.

Nouwen, J. L., Van Belkum, A. & Verbrugh, H. A. 2001. Determinants Of Staphylococcus Aureus Nasal Carriage. *Neth J Med,* 59**,** 126-33.

Nunes, M. C., Shiri, T., Van Niekerk, N., Cutland, C. L., Groome, M. J., Koen, A., Von Gottberg, A., De Gouveia, L., Klugman, K. P., Adrian, P. V. & Madhi, S. A. 2013. Acquisition Of Streptococcus Pneumoniae In Pneumococcal Conjugate Vaccine-Naive South African Children And Their Mothers. *Pediatr Infect Dis J,* 32**,** E192-205.

Nunes, M. C., Von Gottberg, A., De Gouveia, L., Cohen, C., Kuwanda, L., Karstaedt, A. S., Klugman, K. P. & Madhi, S. A. 2011a. Persistent High Burden Of Invasive Pneumococcal Disease In South African Hiv-Infected

Adults In The Era Of An Antiretroviral Treatment Program. *Plos One,* 6**,** E27929.

Nunes, M. C., Von Gottberg, A., De Gouveia, L., Cohen, C., Moore, D. P., Klugman, K. P. & Madhi, S. A. 2011b. The Impact Of Antiretroviral Treatment On The Burden Of Invasive Pneumococcal Disease In South African Children: A Time Series Analysis. *Aids,* 25**,** 453-62.

Nunn, A. J., Mwaba, P., Chintu, C., Mwinga, A., Darbyshire, J. H., Zumla, A. & Collaboration, U.-U. P. L. 2008. Role Of Co-Trimoxazole Prophylaxis In Reducing Mortality In Hiv Infected Adults Being Treated For Tuberculosis: Randomised Clinical Trial. *Bmj,* 337**,** A257.

Nuorti, J. P., Butler, J. C., Gelling, L., Kool, J. L., Reingold, A. L. & Vugia, D. J. 2000. Epidemiologic Relation Between Hiv And Invasive Pneumococcal Disease In San Francisco County, California. *Ann Intern Med,* 132**,** 182-90.

Nurkka, A., Ahman, H., Yaich, M., Eskola, J. & Kayhty, H. 2001. Serum And Salivary Anti-Capsular Antibodies In Infants And Children Vaccinated With Octavalent Pneumococcal Conjugate Vaccines, Pncd And Pnct. *Vaccine,* 20**,** 194-201.

Nyren, P. 1987. Enzymatic Method For Continuous Monitoring Of Dna Polymerase Activity. *Analytical Biochemistry,* 167**,** 235-8.

Nzenze, S. A., Shiri, T., Nunes, M. C., Klugman, K. P., Kahn, K., Twine, R., De Gouveia, L., Von Gottberg, A. & Madhi, S. A. 2013. Temporal Changes In Pneumococcal Colonization In A Rural African Community With High Hiv

Prevalence Following Routine Infant Pneumococcal Immunization. *Pediatr Infect Dis J,* 32**,** 1270-8.

O'brien, K. L., Bronsdon, M. A., Dagan, R., Yagupsky, P., Janco, J., Elliott, J., Whitney, C. G., Yang, Y. H., Robinson, L. G., Schwartz, B. & Carlone, G. M. 2001. Evaluation Of A Medium (Stgg) For Transport And Optimal Recovery Of Streptococcus Pneumoniae From Nasopharyngeal Secretions Collected During Field Studies. *Journal Of Clinical Microbiology,* 39**,** 1021-4.

O'brien, K. L., Millar, E. V., Zell, E. R., Bronsdon, M., Weatherholtz, R., Reid, R., Becenti, J., Kvamme, S., Whitney, C. G. & Santosham, M. 2007. Effect Of Pneumococcal Conjugate Vaccine On Nasopharyngeal Colonization Among Immunized And Unimmunized Children In A Community-Randomized Trial. *J Infect Dis,* 196**,** 1211-20.

O'brien, K. L. & Nohynek, H. 2003. Report From A Who Working Group: Standard Method For Detecting Upper Respiratory Carriage Of Streptococcus Pneumoniae. *The Pediatric Infectious Disease Journal,* 22**,** E1-11.

O'brien, K. L., Wolfson, L. J., Watt, J. P., Henkle, E., Deloria-Knoll, M., Mccall, N., Lee, E., Mulholland, K., Levine, O. S. & Cherian, T. 2009. Burden Of Disease Caused By Streptococcus Pneumoniae In Children Younger Than 5 Years: Global Estimates. *Lancet,* 374**,** 893-902.

Obaro, S. & Adegbola, R. 2002. The Pneumococcus: Carriage, Disease And Conjugate Vaccines. *Journal Of Medical Microbiology,* 51**,** 98-104.

Obaro, S. K., Adegbola, R. A., Banya, W. A. & Greenwood, B. M. 1996. Carriage Of Pneumococci After Pneumococcal Vaccination. *Lancet,* 348**,** 271-2.

Ochs, M. M., Bartlett, W., Briles, D. E., Hicks, B., Jurkuvenas, A., Lau, P., Ren, B. & Millar, A. 2008. Vaccine-Induced Human Antibodies To Pspa Augment Complement C3 Deposition On Streptococcus Pneumoniae. *Microb Pathog,* 44**,** 204-14.

Ogunniyi, A. D., Grabowicz, M., Briles, D. E., Cook, J. & Paton, J. C. 2007a. Development Of A Vaccine Against Invasive Pneumococcal Disease Based On Combinations Of Virulence Proteins Of Streptococcus Pneumoniae. *Infect Immun,* 75**,** 350-7.

Ogunniyi, A. D., Grabowicz, M., Mahdi, L. K., Cook, J., Gordon, D. L., Sadlon, T. A. & Paton, J. C. 2009. Pneumococcal Histidine Triad Proteins Are Regulated By The Zn2+-Dependent Repressor Adcr And Inhibit Complement Deposition Through The Recruitment Of Complement Factor H. *Faseb J,* 23**,** 731-8.

Ogunniyi, A. D., Lemessurier, K. S., Graham, R. M., Watt, J. M., Briles, D. E., Stroeher, U. H. & Paton, J. C. 2007b. Contributions Of Pneumolysin, Pneumococcal Surface Protein A (Pspa), And Pspc To Pathogenicity Of Streptococcus Pneumoniae D39 In A Mouse Model. *Infection And Immunity,* 75**,** 1843-51.

Ogunniyi, A. D., Woodrow, M. C., Poolman, J. T. & Paton, J. C. 2001. Protection Against Streptococcus Pneumoniae Elicited By Immunization With Pneumolysin And Cbpa. *Infect Immun,* 69**,** 5997-6003.

Oh, J., Conlan, S., Polley, E. C., Segre, J. A. & Kong, H. H. 2012. Shifts In Human Skin And Nares Microbiota Of Healthy Children And Adults. *Genome Med,* 4**,** 77.

Oliveira, M. L., Areas, A. P., Campos, I. B., Monedero, V., Perez-Martinez, G., Miyaji, E. N., Leite, L. C., Aires, K. A. & Lee Ho, P. 2006. Induction Of Systemic And Mucosal Immune Response And Decrease In Streptococcus Pneumoniae Colonization By Nasal Inoculation Of Mice With Recombinant Lactic Acid Bacteria Expressing Pneumococcal Surface Antigen A. *Microbes Infect,* 8**,** 1016-24.

Oliver, D. 2003. Microbes and you normal flora. Available: http://www.scq.ubc.ca/microbes-and-you-normal-flora [Accessed March 3 2010].

Oliver, M. B., Van Der Linden, M. P., Kuntzel, S. A., Saad, J. S. & Nahm, M. H. 2013. Discovery Of Streptococcus Pneumoniae Serotype 6 Variants With Glycosyltransferases Synthesizing Two Differing Repeating Units. *The Journal Of Biological Chemistry,* 288**,** 25976-85.

Olsen, G. J., Lane, D. J., Giovannoni, S. J., Pace, N. R. & Stahl, D. A. 1986. Microbial Ecology And Evolution: A Ribosomal Rna Approach. *Annual Review Of Microbiology,* 40**,** 337-65.

Pai, R., Gertz, R. E. & Beall, B. 2006. Sequential Multiplex Pcr Approach For Determining Capsular Serotypes Of Streptococcus Pneumoniae Isolates. *J Clin Microbiol,* 44**,** 124-31.

Pan, E. S., Diep, B. A., Charlebois, E. D., Auerswald, C., Carleton, H. A., Sensabaugh, G. F. & Perdreau-Remington, F. 2005. Population Dynamics Of Nasal Strains Of Methicillin-Resistant Staphylococcus Aureus--And Their Relation To Community-Associated Disease Activity. *J Infect Dis,* 192**,** 811-8.

Pancholi, V., Fontan, P. & Jin, H. 2003. Plasminogen-Mediated Group A Streptococcal Adherence To And Pericellular Invasion Of Human Pharyngeal Cells. *Microbial Pathogenesis,* 35**,** 293-303.

Parameswaran, P., Jalili, R., Tao, L., Shokralla, S., Gharizadeh, B., Ronaghi, M. & Fire, A. Z. 2007. A Pyrosequencing-Tailored Nucleotide Barcode Design Unveils Opportunities For Large-Scale Sample Multiplexing. *Nucleic Acids Res,* 35**,** E130.

Park, I. H., Pritchard, D. G., Cartee, R., Brandao, A., Brandileone, M. C. & Nahm, M. H. 2007. Discovery Of A New Capsular Serotype (6c) Within Serogroup 6 Of Streptococcus Pneumoniae. *J Clin Microbiol,* 45**,** 1225-33.

Parry, C. M., Diep, T. S., Wain, J., Hoa, N. T., Gainsborough, M., Nga, D., Davies, C., Phu, N. H., Hien, T. T., White, N. J. & Farrar, J. J. 2000. Nasal Carriage In Vietnamese Children Of Streptococcus Pneumoniae Resistant To Multiple Antimicrobial Agents. *Antimicrob Agents Chemother,* 44**,** 484-8.

Paterson, G. K., Nieminen, L., Jefferies, J. M. & Mitchell, T. J. 2008. Pcla, A Pneumococcal Collagen-Like Protein With Selected Strain Distribution, Contributes To Adherence And Invasion Of Host Cells. *Fems Microbiol Lett,* 285**,** 170-6.

Pearce, B. J., Iannelli, F. & Pozzi, G. 2002. Construction Of New Unencapsulated (Rough) Strains Of Streptococcus Pneumoniae. *Research In Microbiology,* 153**,** 243-7.

Peltola, V. T., Boyd, K. L., Mcauley, J. L., Rehg, J. E. & Mccullers, J. A. 2006. Bacterial Sinusitis And Otitis Media Following Influenza Virus Infection In Ferrets. *Infection And Immunity,* 74**,** 2562-7.

Peltola, V. T. & Mccullers, J. A. 2004. Respiratory Viruses Predisposing To Bacterial Infections: Role Of Neuraminidase. *Pediatr Infect Dis J,* 23**,** S87-97.

Pericone, C. D., Overweg, K., Hermans, P. W. & Weiser, J. N. 2000. Inhibitory And Bactericidal Effects Of Hydrogen Peroxide Production By Streptococcus Pneumoniae On Other Inhabitants Of The Upper Respiratory Tract. *Infection And Immunity,* 68**,** 3990-7.

Petrosino, J. F., Highlander, S., Luna, R. A., Gibbs, R. A. & Versalovic, J. 2009. Metagenomic Pyrosequencing And Microbial Identification. *Clinical Chemistry,* 55**,** 856-66.

Petti, C. A., Polage, C. R. & Schreckenberger, P. 2005. The Role Of 16s Rrna Gene Sequencing In Identification Of Microorganisms Misidentified By Conventional Methods. *Journal Of Clinical Microbiology,* 43**,** 6123-5.

Pettigrew, M. M., Gent, J. F., Revai, K., Patel, J. A. & Chonmaitree, T. 2008. Microbial Interactions During Upper Respiratory Tract Infections. *Emerg Infect Dis,* 14**,** 1584-91.

Phalipon, A. & Corthesy, B. 2003. Novel Functions Of The Polymeric Ig Receptor: Well Beyond Transport Of Immunoglobulins. *Trends Immunol,* 24**,** 55-8.

Picard, C., Puel, A., Bustamante, J., Ku, C. L. & Casanova, J. L. 2003. Primary Immunodeficiencies Associated With Pneumococcal Disease. *Current Opinion In Allergy And Clinical Immunology,* 3**,** 451-9.

Pido-Lopez, J., Kwok, W. W., Mitchell, T. J., Heyderman, R. S. & Williams, N. A. 2011. Acquisition Of Pneumococci Specific Effector And Regulatory Cd4+ T Cells Localising Within Human Upper Respiratory-Tract Mucosal Lymphoid Tissue. *Plos Pathog,* 7**,** E1002396.

Pilishvili, T., Lexau, C., Farley, M. M., Hadler, J., Harrison, L. H., Bennett, N. M., Reingold, A., Thomas, A., Schaffner, W., Craig, A. S., Smith, P. J., Beall, B. W., Whitney, C. G., Moore, M. R. & Network, A. B. C. S. E. I. P. 2010. Sustained Reductions In Invasive Pneumococcal Disease In The Era Of Conjugate Vaccine. *J Infect Dis,* 201**,** 32-41.

Pimenta, F. C., Miyaji, E. N., Areas, A. P., Oliveira, M. L., De Andrade, A. L., Ho, P. L., Hollingshead, S. K. & Leite, L. C. 2006. Intranasal Immunization With The Cholera Toxin B Subunit-Pneumococcal Surface Antigen A Fusion Protein Induces Protection Against Colonization With Streptococcus Pneumoniae And Has Negligible Impact On The Nasopharyngeal And Oral Microbiota Of Mice. *Infect Immun,* 74**,** 4939-44.

Pimenta, F. C., Roundtree, A., Soysal, A., Bakir, M., Du Plessis, M., Wolter, N., Von Gottberg, A., Mcgee, L., Carvalho Mda, G. & Beall, B. 2013. Sequential Triplex Real-Time Pcr Assay For Detecting 21 Pneumococcal Capsular

Serotypes That Account For A High Global Disease Burden. *J Clin Microbiol,* 51**,** 647-52.

Pletz, M. W., Maus, U., Krug, N., Welte, T. & Lode, H. 2008. Pneumococcal Vaccines: Mechanism Of Action, Impact On Epidemiology And Adaption Of The Species. *Int J Antimicrob Agents,* 32**,** 199-206.

Poehling, K. A., Talbot, T. R., Griffin, M. R., Craig, A. S., Whitney, C. G., Zell, E., Lexau, C. A., Thomas, A. R., Harrison, L. H., Reingold, A. L., Hadler, J. L., Farley, M. M., Anderson, B. J. & Schaffner, W. 2006. Invasive Pneumococcal Disease Among Infants Before And After Introduction Of Pneumococcal Conjugate Vaccine. *Jama,* 295**,** 1668-74.

Polack, F. P., Flayhart, D. C., Zahurak, M. L., Dick, J. D. & Willoughby, R. E. 2000. Colonization By Streptococcus Penumoniae In Human Immunodeficiency Virus-Infected Children. *Pediatr Infect Dis J,* 19**,** 608-12.

Porat, N., Greenberg, D., Givon-Lavi, N., Shuval, D. S., Trefler, R., Segev, O., Hanage, W. P. & Dagan, R. 2006. The Important Role Of Nontypable Streptococcus Pneumoniae International Clones In Acute Conjunctivitis. *J Infect Dis,* 194**,** 689-96.

Post, J. C., Stoodley, P., Hall-Stoodley, L. & Ehrlich, G. D. 2004. The Role Of Biofilms In Otolaryngologic Infections. *Curr Opin Otolaryngol Head Neck Surg,* 12**,** 185-90.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P.,

Bertalan, M., Batto, J. M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H. B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Jian, M., Zhou, Y., Li, Y., Zhang, X., Qin, N., Yang, H., Wang, J., Brunak, S., Dore, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Bork, P. & Ehrlich, S. D. 2010. A Human Gut Microbial Gene Catalogue Established By Metagenomic Sequencing. *Nature,* 464**,** 59-65.

Qiu, X., Wu, L., Huang, H., Mcdonel, P. E., Palumbo, A. V., Tiedje, J. M. & Zhou, J. 2001. Evaluation Of Pcr-Generated Chimeras, Mutations, And Heteroduplexes With 16s Rrna Gene-Based Cloning. *Appl Environ Microbiol,* 67**,** 880-7.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. & Glockner, F. O. 2013. The Silva Ribosomal Rna Gene Database Project: Improved Data Processing And Web-Based Tools. *Nucleic Acids Res,* 41**,** D590-6.

Quin, L. R., Moore, Q. C., 3rd & Mcdaniel, L. S. 2007. Pneumolysin, Pspa, And Pspc Contribute To Pneumococcal Evasion Of Early Innate Immune Responses During Bacteremia In Mice. *Infection And Immunity,* 75**,** 2067-70.

Quince, C., Lanzen, A., Curtis, T. P., Davenport, R. J., Hall, N., Head, I. M., Read, L. F. & Sloan, W. T. 2009. Accurate Determination Of Microbial Diversity From 454 Pyrosequencing Data. *Nature Methods,* 6**,** 639-41.

Quince, C., Lanzen, A., Davenport, R. J. & Turnbaugh, P. J. 2011. Removing Noise From Pyrosequenced Amplicons. *Bmc Bioinformatics,* 12**,** 38.

Quinlan, A. R., Stewart, D. A., Stromberg, M. P. & Marth, G. T. 2008. Pyrobayes: An Improved Base Caller For Snp Discovery In Pyrosequences. *Nat Methods,* 5**,** 179-81.

Raphael, G. D., Jeney, E. V., Baraniuk, J. N., Kim, I., Meredith, S. D. & Kaliner, M. A. 1989. Pathophysiology Of Rhinitis. Lactoferrin And Lysozyme In Nasal Secretions. *The Journal Of Clinical Investigation,* 84**,** 1528-35.

Rapola, S., Salo, E., Kiiski, P., Leinonen, M. & Takala, A. K. 1997. Comparison Of Four Different Sampling Methods For Detecting Pharyngeal Carriage Of Streptococcus Pneumoniae And Haemophilus Influenzae In Children. *J Clin Microbiol,* 35**,** 1077-9.

Reeder, J. & Knight, R. 2010. Rapidly Denoising Pyrosequencing Amplicon Reads By Exploiting Rank-Abundance Distributions. *Nature Methods,* 7**,** 668-9.

Regev-Yochay, G., Dagan, R., Raz, M., Carmeli, Y., Shainberg, B., Derazne, E., Rahav, G. & Rubinstein, E. 2004a. Association Between Carriage Of Streptococcus Pneumoniae And Staphylococcus Aureus In Children. *Jama,* 292**,** 716-20.

Regev-Yochay, G., Hanage, W. P., Trzcinski, K., Rifas-Shiman, S. L., Lee, G., Bessolo, A., Huang, S. S., Pelton, S. I., Mcadam, A. J., Finkelstein, J. A., Lipsitch, M. & Malley, R. 2010. Re-Emergence Of The Type 1 Pilus Among Streptococcus Pneumoniae Isolates In Massachusetts, Usa. *Vaccine,* 28**,** 4842-6.

Regev-Yochay, G., Raz, M., Dagan, R., Porat, N., Shainberg, B., Pinco, E., Keller, N. & Rubinstein, E. 2004b. Nasopharyngeal Carriage Of Streptococcus

Pneumoniae By Adults And Children In Community And Family Settings. *Clin Infect Dis,* 38**,** 632-9.

Regev-Yochay, G., Trzcinski, K., Thompson, C. M., Malley, R. & Lipsitch, M. 2006. Interference Between Streptococcus Pneumoniae And Staphylococcus Aureus: In Vitro Hydrogen Peroxide-Mediated Killing By Streptococcus Pneumoniae. *Journal Of Bacteriology,* 188**,** 4996-5001.

Reichmann, P., Konig, A., Marton, A. & Hakenbeck, R. 1996. Penicillin-Binding Proteins As Resistance Determinants In Clinical Isolates Of Streptococcus Pneumoniae. *Microb Drug Resist,* 2**,** 177-81.

Reisman, J., Rudolph, K., Bruden, D., Hurlburt, D., Bruce, M. G. & Hennessy, T. 2013. Risk Factors For Pneumococcal Colonization Of The Nasopharynx In Alaska Native Adults And Children. *Journal Of The Pediatric Infectious Diseases Society* 1-8.

Ren, B., Li, J., Genschmer, K., Hollingshead, S. K. & Briles, D. E. 2012. The Absence Of Pspa Or Presence Of Antibody To Pspa Facilitates The Complement-Dependent Phagocytosis Of Pneumococci In Vitro. *Clin Vaccine Immunol,* 19**,** 1574-82.

Ricketson, L. J., Wood, M. L., Vanderkooi, O. G., Macdonald, J. C., Martin, I. E., Demczuk, W. H. B. & Kellner, J. D. 2014. Trends In Asymptomatic Nasopharyngeal Colonization With Streptococcus Pneumoniae After Introduction Of The 13-Valent Pneumococcal Conjugate Vaccine In Calgary, Canada. *The Pediatric Infectious Disease Journal,* 33**,** 724-30.

Ringel-Kulka, T., Cheng, J., Ringel, Y., Salojarvi, J., Carroll, I., Palva, A., De Vos, W. M. & Satokari, R. 2013. Intestinal Microbiota In Healthy U.S. Young Children And Adults--A High Throughput Microarray Analysis. *Plos One,* 8**,** E64315.

Rioux, S., Neyt, C., Di Paolo, E., Turpin, L., Charland, N., Labbe, S., Mortier, M. C., Mitchell, T. J., Feron, C., Martin, D. & Poolman, J. T. 2011. Transcriptional Regulation, Occurrence And Putative Role Of The Pht Family Of Streptococcus Pneumoniae. *Microbiology,* 157**,** 336-48.

Rivera-Olivero, I. A., Blommaart, M., Bogaert, D., Hermans, P. W. & De Waard, J. H. 2009. Multiplex Pcr Reveals A High Rate Of Nasopharyngeal Pneumococcal 7-Valent Conjugate Vaccine Serotypes Co-Colonizing Indigenous Warao Children In Venezuela. *J Med Microbiol,* 58**,** 584-7.

Roca, A., Hill, P. C., Townend, J., Egere, U., Antonio, M., Bojang, A., Akisanya, A., Litchfield, T., Nsekpong, D. E., Oluwalana, C., Howie, S. R., Greenwood, B. & Adegbola, R. A. 2011. Effects Of Community-Wide Vaccination With Pcv-7 On Pneumococcal Nasopharyngeal Carriage In The Gambia: A Cluster-Randomized Trial. *Plos Med,* 8**,** E1001107.

Rodgers, G. L., Arguedas, A., Cohen, R. & Dagan, R. 2009. Global Serotype Distribution Among Streptococcus Pneumoniae Isolates Causing Otitis Media In Children: Potential Implications For Pneumococcal Conjugate Vaccines. *Vaccine,* 27**,** 3802-10.

Rodrigues, F., Foster, D., Nicoli, E., Trotter, C., Vipond, B., Muir, P., Goncalves, G., Januario, L. & Finn, A. 2013a. Relationships Between Rhinitis Symptoms,

Respiratory Viral Infections And Nasopharyngeal Colonization With Streptococcus Pneumoniae, Haemophilus Influenzae And Staphylococcus Aureus In Children Attending Daycare. *Pediatr Infect Dis J,* 32**,** 227-32.

Rodrigues, F., Morales-Aza, B., Turner, K. M., Sikora, P., Gould, K., Hinds, J., Goncalves, G., Januario, L. & Finn, A. 2013b. Multiple Streptococcus Pneumoniae Serotypes In Aural Discharge Samples From Children With Acute Otitis Media With Spontaneous Otorrhea. *J Clin Microbiol,* 51**,** 3409-11.

Rodriguez-Barradas, M. C., Tharapel, R. A., Groover, J. E., Giron, K. P., Lacke, C. E., Houston, E. D., Hamill, R. J., Steinhoff, M. C. & Musher, D. M. 1997. Colonization By Streptococcus Pneumoniae Among Human Immunodeficiency Virus-Infected Adults: Prevalence Of Antibiotic Resistance, Impact Of Immunization, And Characterization By Polymerase Chain Reaction With Box Primers Of Isolates From Persistent S. Pneumoniae Carriers. *J Infect Dis,* 175**,** 590-7.

Roh, S. W., Abell, G. C., Kim, K. H., Nam, Y. D. & Bae, J. W. 2010. Comparing Microarrays And Next-Generation Sequencing Technologies For Microbial Ecology Research. *Trends Biotechnol,* 28**,** 291-9.

Rolo, D., Ardanuy, C., Fleites, A., Martin, R. & Linares, J. 2009. Diversity Of Pneumococcal Surface Protein A (Pspa) Among Prevalent Clones In Spain. *Bmc Microbiol,* 9**,** 80.

Rosen, C., Christensen, P., Hovelius, B. & Prellner, K. 1984. A Longitudinal Study Of The Nasopharyngeal Carriage Of Pneumococci As Related To

Pneumococcal Vaccination In Children Attending Day-Care Centres. *Acta Oto-Laryngologica,* 98**,** 524-32.

Rosenow, C., Ryan, P., Weiser, J. N., Johnson, S., Fontan, P., Ortqvist, A. & Masure, H. R. 1997. Contribution Of Novel Choline-Binding Proteins To Adherence, Colonization And Immunogenicity Of Streptococcus Pneumoniae. *Mol Microbiol,* 25**,** 819-29.

Rosenthal, R. S., Folkening, W. J., Miller, D. R. & Swim, S. C. 1983. Resistance Of O-Acetylated Gonococcal Peptidoglycan To Human Peptidoglycan-Degrading Enzymes. *Infection And Immunity,* 40**,** 903-11.

Rothberg, J. M. & Leamon, J. H. 2008. The Development And Impact Of 454 Sequencing. *Nature Biotechnology,* 26**,** 1117-24.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. & Barrell, B. 2000. Artemis: Sequence Visualization And Annotation. *Bioinformatics,* 16**,** 944-5.

Rychlik, W. 1995. Selection Of Primers For Polymerase Chain Reaction. *Molecular Biotechnology,* 3**,** 129-34.

Salter, S., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., Turner, P., Parkhill, J., Loman, N. & Walker, A. W. 2014. Reagent Contamination Can Critically Impact Sequence-Based Microbiome Analysis. *Biorxiv* [Online]. Available: http://dx.doi.org/10.1101/007187 [Accessed 25 July 2014]

Salter, S. J., Hinds, J., Gould, K. A., Lambertsen, L., Hanage, W. P., Antonio, M., Turner, P., Hermans, P. W., Bootsma, H. J., O'brien, K. L. & Bentley, S. D. 2012. Variation At The Capsule Locus, Cps, Of Mistyped And Non-Typable Streptococcus Pneumoniae Isolates. *Microbiology,* 158**,** 1560-9.

Sampo, M., Ghazouani, O., Cadiou, D., Trichet, E., Hoffart, L. & Drancourt, M. 2013. Dolosigranulum Pigrum Keratitis: A Three-Case Series. *Bmc Ophthalmol,* 13**,** 31.

Sanger, F., Nicklen, S. & Coulson, A. R. 1977. Dna Sequencing With Chain-Terminating Inhibitors. *Proceedings Of The National Academy Of Sciences Of The United States Of America,* 74**,** 5463-7.

Satzke, C., Turner, P., Virolainen-Julkunen, A., Adria, P. V., Antonio, M., Hare, K. M., A.M., H.-R., Leach, A. J., Klugman, K. P., Porter, B. D., Sa-Leao, R., Scott, J. A., Nohynek, H. & O'brien, K. L. 2014. Standard Method For Detecting Upper Respiratory Carriage Of Streptococcus Pneumoniae: Updated Recommendations From The World Health Organization Pneumococcal Carriage Working Group. *Vaccine,* 32**,** 165-179.

Saxena, D., Li, Y., Yang, L., Pei, Z., Poles, M., Abrams, W. R. & Malamud, D. 2012. Human Microbiome And Hiv/Aids. *Current Hiv/Aids Reports,* 9**,** 44-51.

Schaad, N. W., Postnikova, E., Sechler, A., Claflin, L. E., Vidaver, A. K., Jones, J. B., Agarkova, I., Ignatov, A., Dickstein, E. & Ramundo, B. A. 2008. Reclassification Of Subspecies Of Acidovorax Avenae As A. Avenae (Manns 1905) Emend., A. Cattleyae (Pavarino, 1911) Comb. Nov., A. Citrulli Schaad

Et Al., 1978) Comb. Nov., And Proposal Of A. Oryzae Sp. Nov. *Syst Appl Microbiol,* 31**,** 434-46.

Schaechter, M., Engleberg, C., Eisenstein, B. I. & Medoff, G. (Eds.) 1998. *Mechanisms Of Microbial Disease*: Lippincott Williams & Wilkins.

Schloss, P. D., Gevers, D. & Westcott, S. L. 2011. Reducing The Effects Of Pcr Amplification And Sequencing Artifacts On 16s Rrna-Based Studies. *Plos One,* 6**,** E27310.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J. & Weber, C. F. 2009. Introducing Mothur: Open-Source, Platform-Independent, Community-Supported Software For Describing And Comparing Microbial Communities. *Applied And Environmental Microbiology,* 75**,** 7537-41.

Scott, J. R., Hanage, W. P., Lipsitch, M., Millar, E. V., Moulton, L. H., Hinds, J., Reid, R., Santosham, M. & O'brien, K. L. 2011. Pneumococcal Sequence Type Replacement Among American Indian Children: A Comparison Of Pre- And Routine-Pcv7 Eras. *Vaccine,* 30**,** 2376-81.

Scott, J. R., Hinds, J., Gould, K. A., Millar, E. V., Reid, R., Santosham, M., O'brien, K. L. & Hanage, W. P. 2012. Nontypeable Pneumococcal Isolates Among Navajo And White Mountain Apache Communities: Are These Really A Cause Of Invasive Disease? *J Infect Dis,* 206**,** 73-80.

Seepersaud, R., Hanniffy, S. B., Mayne, P., Sizer, P., Le Page, R. & Wells, J. M. 2005. Characterization Of A Novel Leucine-Rich Repeat Protein Antigen

From Group B Streptococci That Elicits Protective Immunity. *Infect Immun,* 73**,** 1671-83.

Senkovich, O., Cook, W. J., Mirza, S., Hollingshead, S. K., Protasevich, Ii, Briles, D. E. & Chattopadhyay, D. 2007. Structure Of A Complex Of Human Lactoferrin N-Lobe With Pneumococcal Surface Protein A Provides Insight Into Microbial Defense Mechanism. *Journal Of Molecular Biology,* 370**,** 701-13.

Shak, J. R., Vidal, J. E. & Klugman, K. P. 2012. Influence Of Bacterial Interactions On Pneumococcal Colonization Of The Nasopharynx. *Trends Microbiol,* 21**,** 129-35.

Shakhnovich, E. A., King, S. J. & Weiser, J. N. 2002. Neuraminidase Expressed By Streptococcus Pneumoniae Desialylates The Lipopolysaccharide Of Neisseria Meningitidis And Haemophilus Influenzae: A Paradigm For Interbacterial Competition Among Pathogens Of The Human Respiratory Tract. *Infect Immun,* 70**,** 7161-4.

Shaper, M., Hollingshead, S. K., Benjamin, W. H., Jr. & Briles, D. E. 2004. Pspa Protects Streptococcus Pneumoniae From Killing By Apolactoferrin, And Antibody To Pspa Enhances Killing Of Pneumococci By Apolactoferrin [Corrected]. *Infection And Immunity,* 72**,** 5031-40.

Shibabaw, A., Abebe, T. & Mihret, A. 2013. Nasal Carriage Rate Of Methicillin Resistant Staphylococcus Aureus Among Dessie Referral Hospital Health Care Workers; Dessie, Northeast Ethiopia. *Antimicrob Resist Infect Control,* 2**,** 25.

Shrestha, R. K., Rosenberg, T., Makarovsky, D., Eckshtain-Levi, N., Zelinger, E., Kopelowitz, J., Sikorski, J. & Burdman, S. 2013. Phenotypic Variation In The Plant Pathogenic Bacterium Acidovorax Citrulli. *Plos One,* 8**,** E73189.

Simell, B., Auranen, K., Kayhty, H., Goldblatt, D., Dagan, R. & O'brien, K. L. 2012a. The Fundamental Link Between Pneumococcal Carriage And Disease. *Expert Rev Vaccines,* 11**,** 841-55.

Simell, B., Kilpi, T. M. & Kayhty, H. 2002. Pneumococcal Carriage And Otitis Media Induce Salivary Antibodies To Pneumococcal Capsular Polysaccharides In Children. *J Infect Dis,* 186**,** 1106-14.

Simell, B., Lahdenkari, M., Reunanen, A., Kayhty, H. & Vakevainen, M. 2008. Effects Of Ageing And Gender On Naturally Acquired Antibodies To Pneumococcal Capsular Polysaccharides And Virulence-Associated Proteins. *Clin Vaccine Immunol,* 15**,** 1391-7.

Simell, B., Nurkka, A., Lahdenkari, M., Givon-Lavi, N., Kayhty, H., Dagan, R. & Jokinen, J. 2012b. Association Of Serotype-Specific Antibody Concentrations And Functional Antibody Titers With Subsequent Pneumococcal Carriage In Toddlers Immunized With A 9-Valent Pneumococcal Conjugate Vaccine. *Clin Vaccine Immunol,* 19**,** 96-9.

Singleton, R. J., Hennessy, T. W., Bulkow, L. R., Hammitt, L. L., Zulz, T., Hurlburt, D. A., Butler, J. C., Rudolph, K. & Parkinson, A. 2007. Invasive Pneumococcal Disease Caused By Nonvaccine Serotypes Among Alaska Native Children With High Levels Of 7-Valent Pneumococcal Conjugate Vaccine Coverage. *Jama,* 297**,** 1784-92.

Sleeman, K. L., Griffiths, D., Shackley, F., Diggle, L., Gupta, S., Maiden, M. C., Moxon, E. R., Crook, D. W. & Peto, T. E. 2006. Capsular Serotype-Specific Attack Rates And Duration Of Carriage Of Streptococcus Pneumoniae In A Population Of Children. *The Journal Of Infectious Diseases,* 194**,** 682-8.

Slotved, H. C., Kaltoft, M., Skovsted, I. C., Kerrn, M. B. & Espersen, F. 2004. Simple, Rapid Latex Agglutination Test For Serotyping Of Pneumococci (Pneumotest-Latex). *Journal Of Clinical Microbiology,* 42**,** 2518-22.

Smith, M. I., Yatsunenko, T., Manary, M. J., Trehan, I., Mkakosya, R., Cheng, J., Kau, A. L., Rich, S. S., Concannon, P., Mychaleckyj, J. C., Liu, J., Houpt, E., Li, J. V., Holmes, E., Nicholson, J., Knights, D., Ursell, L. K., Knight, R. & Gordon, J. I. 2013. Gut Microbiomes Of Malawian Twin Pairs Discordant For Kwashiorkor. *Science,* 339**,** 548-54.

Sohn, J. H., Kwon, K. K., Kang, J. H., Jung, H. B. & Kim, S. J. 2004. Novosphingobium Pentaromativorans Sp. Nov., A High-Molecular-Mass Polycyclic Aromatic Hydrocarbon-Degrading Bacterium Isolated From Estuarine Sediment. *Int J Syst Evol Microbiol,* 54**,** 1483-7.

Soltis, P. S. & Soltis, D. E. 2003. Applying The Bootstrap In Phylogeny Reconstruction. *Statistical Science,* 18**,** 256-267.

Sorensen, U. B., Henrichsen, J., Chen, H. C. & Szu, S. C. 1990. Covalent Linkage Between The Capsular Polysaccharide And The Cell Wall Peptidoglycan Of Streptococcus Pneumoniae Revealed By Immunochemical Methods. *Microb Pathog,* 8**,** 325-34.

Spika, J. S., Facklam, R. R., Plikaytis, B. D. & Oxtoby, M. J. 1991. Antimicrobial Resistance Of Streptococcus Pneumoniae In The United States, 1979-1987. The Pneumococcal Surveillance Working Group. *J Infect Dis,* 163**,** 1273-8.

Stamatakis, A., Ludwig, T. & Meier, H. 2005. Raxml-Iii: A Fast Program For Maximum Likelihood-Based Inference Of Large Phylogenetic Trees. *Bioinformatics,* 21**,** 456-63.

Stein, K. E. 1992. Thymus-Independent And Thymus-Dependent Responses To Polysaccharide Antigens. *J Infect Dis,* 165 Suppl 1**,** S49-52.

Sullivan, M. J., Petty, N. K. & Beatson, S. A. 2011. Easyfig: A Genome Comparison Visualizer. *Bioinformatics,* 27**,** 1009-10.

Sun, K., Johansen, F. E., Eckmann, L. & Metzger, D. W. 2004. An Important Role For Polymeric Ig Receptor-Mediated Transport Of Iga In Protection Against Streptococcus Pneumoniae Nasopharyngeal Carriage. *J Immunol,* 173**,** 4576-81.

Syrjanen, R. K., Kilpi, T. M., Kaijalainen, T. H., Herva, E. E. & Takala, A. K. 2001. Nasopharyngeal Carriage Of Streptococcus Pneumoniae In Finnish Children Younger Than 2 Years Old. *J Infect Dis,* 184**,** 451-9.

Takeshita, T., Suzuki, N., Nakano, Y., Yasui, M., Yoneda, M., Shimazaki, Y., Hirofuji, T. & Yamashita, Y. 2012. Discrimination Of The Oral Microbiota Associated With High Hydrogen Sulfide And Methyl Mercaptan Production. *Sci Rep,* 2**,** 215.

Talkington, D. F., Brown, B. G., Tharpe, J. A., Koenig, A. & Russell, H. 1996. Protection Of Mice Against Fatal Pneumococcal Challenge By Immunization With Pneumococcal Surface Adhesin A (Psaa). *Microb Pathog,* 21**,** 17-22.

Tan, T. T., Morgelin, M., Forsgren, A. & Riesbeck, K. 2007. Haemophilus Influenzae Survival During Complement-Mediated Attacks Is Promoted By Moraxella Catarrhalis Outer Membrane Vesicles. *The Journal Of Infectious Diseases,* 195**,** 1661-70.

Tanaka, N., Fukuyama, S., Fukuiwa, T., Kawabata, M., Sagara, Y., Ito, H. O., Miwa, Y., Nagatake, T., Kiyono, H. & Kurono, Y. 2007. Intranasal Immunization With Phosphorylcholine Induces Antigen Specific Mucosal And Systemic Immune Responses In Mice. *Vaccine,* 25**,** 2680-7.

Tancrede, C. 1992. Role Of Human Microflora In Health And Disease. *Eur J Clin Microbiol Infect Dis,* 11**,** 1012-5.

Tanner, M. A., Goebel, B. M., Dojka, M. A. & Pace, N. R. 1998. Specific Ribosomal Dna Sequences From Diverse Environmental Settings Correlate With Experimental Contaminants. *Appl Environ Microbiol,* 64**,** 3110-3.

Temime, L., Boelle, P. Y., Opatowski, L. & Guillemot, D. 2008. Impact Of Capsular Switch On Invasive Pneumococcal Disease Incidence In A Vaccinated Population. *Plos One,* 3**,** E3244.

Tettelin, H., Nelson, K. E., Paulsen, I. T., Eisen, J. A., Read, T. D., Peterson, S., Heidelberg, J., Deboy, R. T., Haft, D. H., Dodson, R. J., Durkin, A. S., Gwinn, M., Kolonay, J. F., Nelson, W. C., Peterson, J. D., Umayam, L. A., White, O., Salzberg, S. L., Lewis, M. R., Radune, D., Holtzapple, E., Khouri,

H., Wolf, A. M., Utterback, T. R., Hansen, C. L., Mcdonald, L. A., Feldblyum, T. V., Angiuoli, S., Dickinson, T., Hickey, E. K., Holt, I. E., Loftus, B. J., Yang, F., Smith, H. O., Venter, J. C., Dougherty, B. A., Morrison, D. A., Hollingshead, S. K. & Fraser, C. M. 2001. Complete Genome Sequence Of A Virulent Isolate Of Streptococcus Pneumoniae. *Science,* 293**,** 498-506.

Thomas, C. M. & Nielsen, K. M. 2005. Mechanisms Of, And Barriers To, Horizontal Gene Transfer Between Bacteria. *Nat Rev Microbiol,* 3**,** 711-21.

Thompson, G. G. & Eithers, P. C. 2003. Effect Of Species Richness And Relative Abundance On The Shape Of The Species Accumulation Curve. *Austral Ecology,* 28**,** 355-360.

Todar, K. 2012. *Todar's Online Textbook Of Bacteriology* [Online]: (http://www.textbookofbacteriology.net/S.pneumoniae.html) [Accessed March 31].

Tong, H. H., James, M., Grants, I., Liu, X., Shi, G. & Demaria, T. F. 2001. Comparison Of Structural Changes Of Cell Surface Carbohydrates In The Eustachian Tube Epithelium Of Chinchillas Infected With A Streptococcus Pneumoniae Neuraminidase-Deficient Mutant Or Its Isogenic Parent Strain. *Microbial Pathogenesis,* 31**,** 309-17.

Tseng, H. J., Mcewan, A. G., Paton, J. C. & Jennings, M. P. 2002. Virulence Of Streptococcus Pneumoniae: Psaa Mutants Are Hypersensitive To Oxidative Stress. *Infect Immun,* 70**,** 1635-9.

Turner, P., Hinds, J., Turner, C., Jankhot, A., Gould, K., Bentley, S. D., Nosten, F. & Goldblatt, D. 2011. Improved Detection Of Nasopharyngeal Cocolonization By Multiple Pneumococcal Serotypes By Use Of Latex Agglutination Or Molecular Serotyping By Microarray. *J Clin Microbiol,* 49**,** 1784-9.

Turner, P., Turner, C., Jankhot, A., Helen, N., Lee, S. J., Day, N. P., White, N. J., Nosten, F. & Goldblatt, D. 2012. A Longitudinal Study Of Streptococcus Pneumoniae Carriage In A Cohort Of Infants And Their Mothers On The Thailand-Myanmar Border. *Plos One,* 7**,** E38271.

Turner, P., Turner, C., Jankhot, A., Phakaudom, K., Nosten, F. & Goldblatt, D. 2013. Field Evaluation Of Culture Plus Latex Sweep Serotyping For Detection Of Multiple Pneumococcal Serotype Colonisation In Infants And Young Children. *Plos One,* 8**,** E67933.

Uehara, Y., Nakama, H., Agematsu, K., Uchida, M., Kawakami, Y., Abdul Fattah, A. S. & Maruchi, N. 2000. Bacterial Interference Among Nasal Inhabitants: Eradication Of Staphylococcus Aureus From Nasal Cavities By Artificial Implantation Of Corynebacterium Sp. *J Hosp Infect,* 44**,** 127-33.

Unaids 2012a. 2012 Global Aids Response Progress Report: Malawi Country Report For 2010 And 2011. *Geneva: Joint United Nations Programme On Hiv/Aids*.

UNAIDS 2012b. Global Aids Response Progress Report: Malawi Country Report For 2010 And 2012.

UNAIDS 2013a. Global Report: Unaids Report On The Global Aids Epidemic 2013.

UNAIDS 2013b. Global Summary Of The Aids Epidemic: Core Epidemiology Slides.

Usuf, E., Bottomley, C., Adegbola, R. A. & Hall, A. 2014. Pneumococcal Carriage In Sub-Saharan Africa-A Systematic Review. *Plos One,* 9**,** E85001.

Valente, C., Hinds, J., Pinto, F., Brugger, S. D., Gould, K., Muhlemann, K., De Lencastre, H. & Sa-Leao, R. 2012. Decrease In Pneumococcal Co-Colonization Following Vaccination With The Seven-Valent Pneumococcal Conjugate Vaccine. *Plos One,* 7**,** E30235.

Valles, X., Flannery, B., Roca, A., Mandomando, I., Sigauque, B., Sanz, S., Schuchat, A., Levine, M., Soriano-Gabarro, M. & Alonso, P. 2006. Serotype Distribution And Antibiotic Susceptibility Of Invasive And Nasopharyngeal Isolates Of Streptococcus Pneumoniae Among Children In Rural Mozambique. *Trop Med Int Health,* 11**,** 358-66.

Van Der Ven, L. T., Van Den Dobbelsteen, G. P., Nagarajah, B., Van Dijken, H., Dortant, P. M., Vos, J. G. & Roholl, P. J. 1999. A New Rat Model Of Otitis Media Caused By Streptococcus Pneumoniae: Conditions And Application In Immunization Protocols. *Infection And Immunity,* 67**,** 6098-103.

Van Gijlswijk, R. P., Talman, E. G., Janssen, P. J., Snoeijers, S. S., Killian, J., Tanke, H. J. & Heetebrij, R. J. 2001. Universal Linkage System: Versatile Nucleic Acid Labeling Technique. *Expert Review Of Molecular Diagnostics,* 1**,** 81-91.

Van Rossum, A. M., Lysenko, E. S. & Weiser, J. N. 2005. Host And Bacterial Factors Contributing To The Clearance Of Colonization By Streptococcus Pneumoniae In A Murine Model. *Infect Immun,* 73**,** 7718-26.

Varon, E., Levy, C., De La Rocque, F., Boucherat, M., Deforche, D., Podglajen, I., Navel, M. & Cohen, R. 2000. Impact Of Antimicrobial Therapy On Nasopharyngeal Carriage Of Streptococcus Pneumoniae, Haemophilus Influenzae, And Branhamella Catarrhalis In Children With Respiratory Tract Infections. *Clin Infect Dis,* 31**,** 477-81.

Verduin, C. M., Hol, C., Fleer, A., Van Dijk, H. & Van Belkum, A. 2002. Moraxella Catarrhalis: From Emerging To Established Pathogen. *Clin Microbiol Rev,* 15**,** 125-44.

Vestrheim, D. F., Hoiby, E. A., Bergsaker, M. R., Ronning, K., Aaberge, I. S. & Caugant, D. A. 2010. Indirect Effect Of Conjugate Pneumococcal Vaccination In A 2+1 Dose Schedule. *Vaccine,* 28**,** 2214-21.

VIMS: Global Vaccine Introduction Status. 2013. Vaccine Information Management System (VIMS). Accelarated Vaccine Initiative March, 2013 Report. John Hopkins Bloomberg School Of Public Health.

Vollmer, W., Blanot, D. & De Pedro, M. A. 2008. Peptidoglycan Structure And Architecture. *Fems Microbiology Reviews,* 32**,** 149-67.

Vu, H. T., Yoshida, L. M., Suzuki, M., Nguyen, H. A., Nguyen, C. D., Nguyen, A. T., Oishi, K., Yamamoto, T., Watanabe, K. & Vu, T. D. 2011. Association Between Nasopharyngeal Load Of Streptococcus Pneumoniae, Viral

Coinfection, And Radiologically Confirmed Pneumonia In Vietnamese Children. *Pediatr Infect Dis J,* 30**,** 11-8.

Vujkovic-Cvijin, I., Dunham, R. M., Iwai, S., Maher, M. C., Albright, R. G., Broadhurst, M. J., Hernandez, R. D., Lederman, M. M., Huang, Y., Somsouk, M., Deeks, S. G., Hunt, P. W., Lynch, S. V. & Mccune, J. M. 2013. Dysbiosis Of The Gut Microbiota Is Associated With Hiv Disease Progression And Tryptophan Catabolism. *Science Translational Medicine,* 5**,** 193ra91.

Walker, A. W., Sanderson, J. D., Churcher, C., Parkes, G. C., Hudspith, B. N., Rayment, N., Brostoff, J., Parkhill, J., Dougan, G. & Petrovska, L. 2011. High-Throughput Clone Library Analysis Of The Mucosa-Associated Microbiota Reveals Dysbiosis And Differences Between Inflamed And Non-Inflamed Regions Of The Intestine In Inflammatory Bowel Disease. *Bmc Microbiol,* 11**,** 7.

Wang, G. C. & Wang, Y. 1996. The Frequency Of Chimeric Molecules As A Consequence Of Pcr Co-Amplification Of 16s Rrna Genes From Different Bacterial Species. *Microbiology,* 142 (Pt 5)**,** 1107-14.

Ward, N. L., Challacombe, J. F., Janssen, P. H., Henrissat, B., Coutinho, P. M., Wu, M., Xie, G., Haft, D. H., Sait, M., Badger, J., Barabote, R. D., Bradley, B., Brettin, T. S., Brinkac, L. M., Bruce, D., Creasy, T., Daugherty, S. C., Davidsen, T. M., Deboy, R. T., Detter, J. C., Dodson, R. J., Durkin, A. S., Ganapathy, A., Gwinn-Giglio, M., Han, C. S., Khouri, H., Kiss, H., Kothari, S. P., Madupu, R., Nelson, K. E., Nelson, W. C., Paulsen, I., Penn, K., Ren, Q., Rosovitz, M. J., Selengut, J. D., Shrivastava, S., Sullivan, S. A., Tapia, R., Thompson, L. S., Watkins, K. L., Yang, Q., Yu, C., Zafar, N., Zhou, L. &

Kuske, C. R. 2009. Three Genomes From The Phylum Acidobacteria Provide Insight Into The Lifestyles Of These Microorganisms In Soils. *Appl Environ Microbiol,* 75**,** 2046-56.

Wasilauskas, B. L. & Hampton, K. D. 1984. An Analysis Of Streptococcus Pneumoniae Identification Using Biochemical And Serological Procedures. *Diagn Microbiol Infect Dis,* 2**,** 301-7.

Weidenmaier, C. & Peschel, A. 2008. Teichoic Acids And Related Cell-Wall Glycopolymers In Gram-Positive Physiology And Host Interactions. *Nat Rev Microbiol,* 6**,** 276-87.

Weinberger, D. M., Bruden, D. T., Grant, L. R., Lipsitch, M., O'brien, K. L., Pelton, S. I., Sanders, E. A. & Feikin, D. R. 2013. Using Pneumococcal Carriage Data To Monitor Postvaccination Changes In Invasive Disease. *Am J Epidemiol,* 178**,** 1488-95.

Weinberger, D. M., Malley, R. & Lipsitch, M. 2011. Serotype Replacement In Disease After Pneumococcal Vaccination. *Lancet,* 378**,** 1962-73.

Weinberger, D. M., Trzcinski, K., Lu, Y. J., Bogaert, D., Brandes, A., Galagan, J., Anderson, P. W., Malley, R. & Lipsitch, M. 2009. Pneumococcal Capsular Polysaccharide Structure Predicts Serotype Prevalence. *Plos Pathog,* 5**,** E1000476.

Weintraub, A. 2003. Immunology Of Bacterial Polysaccharide Antigens. *Carbohydr Res,* 338**,** 2539-47.

Weiser, J. N., Austrian, R., Sreenivasan, P. K. & Masure, H. R. 1994. Phase Variation In Pneumococcal Opacity: Relationship Between Colonial Morphology And Nasopharyngeal Colonization. *Infection And Immunity,* 62**,** 2582-9.

Weiser, J. N., Bae, D., Fasching, C., Scamurra, R. W., Ratner, A. J. & Janoff, E. N. 2003. Antibody-Enhanced Pneumococcal Adherence Requires Iga1 Protease. *Proceedings Of The National Academy Of Sciences Of The United States Of America,* 100**,** 4215-20.

Weiser, J. N., Shchepetov, M. & Chong, S. T. 1997. Decoration Of Lipopolysaccharide With Phosphorylcholine: A Phase-Variable Characteristic Of Haemophilus Influenzae. *Infect Immun,* 65**,** 943-50.

Weon, H. Y., Anandham, R., Kim, B. Y., Hong, S. B., Jeon, Y. A. & Kwon, S. W. 2009. Dyella Soli Sp. Nov. And Dyella Terrae Sp. Nov., Isolated From Soil. *Int J Syst Evol Microbiol,* 59**,** 1685-90.

Werner, G., Hildebrandt, B. & Witte, W. 2001. Aminoglycoside-Streptothricin Resistance Gene Cluster Aade-Sat4-Apha-3 Disseminated Among Multiresistant Isolates Of Enterococcus Faecium. *Antimicrobial Agents And Chemotherapy,* 45**,** 3267-9.

White, J. R., Nagarajan, N. & Pop, M. 2009. Statistical Methods For Detecting Differentially Abundant Features In Clinical Metagenomic Samples. *Plos Comput Biol,* 5**,** E1000352.

Whitman, W. B., Coleman, D. C. & Wiebe, W. J. 1998. Prokaryotes: The Unseen Majority. *Proc Natl Acad Sci U S A,* 95**,** 6578-83.

WHO. 2007. Pneumococcal Conjugate Vaccine For Childhood Immunization--WHO Position Paper. *Wkly Epidemiol Rec,* 82**,** 93-104.

WHO. 2008. 23-Valent Pneumococcal Polysaccharide Vaccine. Who Position Paper. *Wkly Epidemiol Rec,* 83**,** 373-84.

Wiegant, J. C., Van Gijlswijk, R. P., Heetebrij, R. J., Bezrookove, V., Raap, A. K. & Tanke, H. J. 1999. Uls: A Versatile Method Of Labeling Nucleic Acids For Fish Based On A Monofunctional Reaction Of Cisplatin Derivatives With Guanine Moieties. *Cytogenet Cell Genet,* 87**,** 47-52.

Williams, K. P., Sobral, B. W. & Dickerman, A. W. 2007. A Robust Species Tree For The Alphaproteobacteria. *J Bacteriol,* 189**,** 4578-86.

Winkelstein, J. A. & Tomasz, A. 1977. Activation Of The Alternative Pathway By Pneumococcal Cell Walls. *Journal Of Immunology,* 118**,** 451-4.

Wolf, B., Rey, L. C., Moreira, L. B., Milatovic, D., Fleer, A., Verhoef, J. & Roord, J. J. 2001. Carriage Of Gram-Negative Bacilli In Young Brazilian Children With Community-Acquired Pneumonia. *Int J Infect Dis,* 5**,** 155-9.

Woolfson, A., Huebner, R., Wasas, A., Chola, S., Godfrey-Faussett, P. & Klugman, K. 1997. Nasopharyngeal Carriage Of Community-Acquired, Antibiotic-Resistant Streptococcus Pneumoniae In A Zambian Paediatric Population. *Bull World Health Organ,* 75**,** 453-62.

Wren, J. D., Kulkarni, A., Joslin, J., Butow, R. A. & Garner, H. R. 2002. Cross-Hybridization On Pcr-Spotted Microarrays. *Ieee Engineering In Medicine*

And Biology Magazine : The Quarterly Magazine Of The Engineering In Medicine & Biology Society, 21**,** 71-5.

Wright, A. K., Ferreira, D. M., Gritzfeld, J. F., Wright, A. D., Armitage, K., Jambo, K. C., Bate, E., El Batrawy, S., Collins, A. & Gordon, S. B. 2012. Human Nasal Challenge With Streptococcus Pneumoniae Is Immunising In The Absence Of Carriage. *Plos Pathog,* 8**,** E1002622.

Wyres, K. L., Lambertsen, L. M., Croucher, N. J., Mcgee, L., Von Gottberg, A., Linares, J., Jacobs, M. R., Kristinsson, K. G., Beall, B. W., Klugman, K. P., Parkhill, J., Hakenbeck, R., Bentley, S. D. & Brueggemann, A. B. 2013. Pneumococcal Capsular Switching: A Historical Perspective. *J Infect Dis,* 207**,** 439-49.

Xu, Q., Almudervar, A., Casey, J. R. & Pichichero, M. E. 2012. Nasopharyngeal Bacterial Interactions In Children. *Emerging Infectious Diseases,* 18**,** 1738-45.

Yang, Z. 1997. Paml: A Program Package For Phylogenetic Analysis By Maximum Likelihood. *Comput Appl Biosci,* 13**,** 555-6.

Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R. N., Anokhin, A. P., Heath, A. C., Warner, B., Reeder, J., Kuczynski, J., Caporaso, J. G., Lozupone, C. A., Lauber, C., Clemente, J. C., Knights, D., Knight, R. & Gordon, J. I. 2012. Human Gut Microbiome Viewed Across Age And Geography. *Nature,* 486**,** 222-7.

Ye, R. W., Wang, T., Bedzyk, L. & Croker, K. M. 2001. Applications Of Dna Microarrays In Microbial Systems. *Journal Of Microbiological Methods,* 47**,** 257-72.

Yeh, S. H., Zangwill, K. M., Lee, H., Chang, S. J., Wong, V. I., Greenberg, D. P. & Ward, J. I. 2003. Heptavalent Pneumococcal Vaccine Conjugated To Outer Membrane Protein Of Neisseria Meningitidis Serogroup B And Nasopharyngeal Carriage Of Streptococcus Pneumoniae In Infants. *Vaccine,* 21**,** 2627-31.

Yoshida, Y., Ganguly, S., Bush, C. A. & Cisar, J. O. 2005. Carbohydrate Engineering Of The Recognition Motifs In Streptococcal Co-Aggregation Receptor Polysaccharides. *Molecular Microbiology,* 58**,** 244-56.

Zapun, A., Contreras-Martel, C. & Vernet, T. 2008. Penicillin-Binding Proteins And Beta-Lactam Resistance. *Fems Microbiol Rev,* 32**,** 361-85.

Zar, H. J. 2004. Pneumonia In Hiv-Infected And Hiv-Uninfected Children In Developing Countries: Epidemiology, Clinical Features, And Management. *Curr Opin Pulm Med,* 10**,** 176-82.

Zerbino, D. R. & Birney, E. 2008. Velvet: Algorithms For De Novo Short Read Assembly Using De Bruijn Graphs. *Genome Res,* 18**,** 821-9.

Zhang, J. R., Idanpaan-Heikkila, I., Fischer, W. & Tuomanen, E. I. 1999. Pneumococcal Licd2 Gene Is Involved In Phosphorylcholine Metabolism. *Mol Microbiol,* 31**,** 1477-88.

Zhang, Q., Arnaoutakis, K., Murdoch, C., Lakshman, R., Race, G., Burkinshaw, R. & Finn, A. 2004. Mucosal Immune Responses To Capsular Pneumococcal Polysaccharides In Immunized Preschool Children And Controls With Similar Nasal Pneumococcal Colonization Rates. *Pediatr Infect Dis J,* 23**,** 307-13.

Zhang, Q., Bagrade, L., Bernatoniene, J., Clarke, E., Paton, J. C., Mitchell, T. J., Nunez, D. A. & Finn, A. 2007. Low Cd4 T Cell Immunity To Pneumolysin Is Associated With Nasopharyngeal Carriage Of Pneumococci In Children. *J Infect Dis,* 195**,** 1194-202.

Zhang, Y., Masi, A. W., Barniak, V., Mountzouros, K., Hostetter, M. K. & Green, B. A. 2001. Recombinant Phpa Protein, A Unique Histidine Motif-Containing Protein From Streptococcus Pneumoniae, Protects Mice Against Intranasal Pneumococcal Challenge. *Infection And Immunity,* 69**,** 3827-36.

Zhou, J., Xia, B., Treves, D. S., Wu, L. Y., Marsh, T. L., O'neill, R. V., Palumbo, A. V. & Tiedje, J. M. 2002. Spatial And Resource Factors Influencing High Microbial Diversity In Soil. *Applied And Environmental Microbiology,* 68**,** 326-34.

Appendices

Appendix 1 (a) Sheep bleeding SOP

MLW.SOP.L.MICRO.035 version 2

Background: Whole blood is used in some microbiological media cultures. This blood is usually obtained from either sheep or horses. Sheep or horse blood is free from antibiotics, which may affect the media performance. Sheep are easily kept and they can be bled as required. Sheep do not cause problems when bleeding unlike other animals.

Purpose: The purpose of this SOP is to describe the procedure to be followed to bleed the sheep at MLW safely to obtain sterile good quality blood.

Responsibilities: Media Technician/Lab Technical Staff

Procedure

Materials A good bleeding bench; Tray; Pair of scissors; Clean scalpel blade; Spirit or 70% alcohol; Blood bag (450ml with sodium citrate anticoagulant); Pair of forceps; Nitrile gloves; Cotton wool

Storage conditions It is advised that sheep blood should be stored in the fridge at the temperature range of between 2-8^{0}C. All flammable liquids should be stored in a flammables cabin.

Test procedure

- Carefully lay the sheep on its left side.
- Using a pair of scissors and scalpel blade, shave the right side of the sheep's neck where the jugular vein is near the shoulder.
- Locate the vein.
- Decontaminate the site using cotton wool and spirit or a 70% alcohol swab. Allow the site to dry.
- Clip the blood bag tubing near the needle using a pair of forceps.
- Carefully press the vein to the side of the shoulder and insert the needle into the vein.
- Unclip a pair of forceps and hold the needle carefully, check for the flow of blood into the bag.
- As the blood flows steadily into the bag, regularly mix the blood with anticoagulant.
- Hold until the bag is filled and once again clip the tubing.
- Carefully remove the needle while holding dry cotton wool near the insertion and press gently on the insertion site with the cotton wool after removal of the needle until blood stops flowing.
- Tie a knot in the blood bag tubing.
- Cut off the needle of the blood bag and dispose of it in a sharps bin.
- Sheep should be only be bled once every three months. Each sheep has a tag number on its ear, and the number of the sheep that has been bled is recorded in a logbook, together with the date. The logbook is kept in the Media room at MLW.
- As soon as bleeding is completed, a gardener should be asked to provide

water and sheep feed to the bled sheep to drink and eat respectively.

- A sick sheep is not supposed to be bled until it fully recovers.

***Quality control*.** It is the responsibility of the sheep bleeder to make sure that blood is not contaminated during and after the bleeding process.

Health and safety

- Follow health and safety guidelines as laid down in the MLW Laboratory Health and Safety Manual
- Ensure all sharps are handled with care, and disposed of according to the waste disposal section of the MLW Health and Safety Manuals.

Appendix 1 (b) Sheep Blood Agar (layered) + Gentamicin (SBG)

MLW.SOP.L.MEDIA.009 version1

Materials. Sheep Blood Agar Base (Oxoid – CM0854)**;** Nutrient Agar Base (Oxoid – CM0003)**;** Sheep Blood (see MLW.SOP.L.MICRO.035for sheep bleeding procedure)**;** Codomycin (Gentamicin) 80µg/2ml vial; Distilled Water**;** Sterile 25ml Pipettes**;** Disposable plastic 90mm Petri dishes.

Procedure Nutrient Agar (bottom) layer. Dissolve 28g of Nutrient Agar base powder in 1L of distilled water; Autoclave at 121^{o}C for 15 minutes; Cool to 50^{o}C in the water bath**;** Aseptically dispense 10ml into sterile Petri dishes and allow to set whilst on a flat, level surface.

Procedure Sheep Blood (top) layer

- Dissolve 40g of Sheep Blood Agar base powder in 1L of distilled water
- Autoclave at 121^{o}C for 15 minutes
- Cool to 50^{o}C in the water bath
- Aseptically add 70ml of Sheep Blood and mix.
- Aseptically add 32µl of Codomycin (from 80µg/2ml vial) giving a final concentration of 2.56µg/L
- Aseptically dispense 10ml into a Petri dish already containing a set 10ml layer of nutrient agar and allow to set whilst on a flat, level surface.

***Storage Conditions*.** Dry all plates in the 37^{o}C aerobic incubator overnight before storing them at 2-8^{o}C until required.

Quality control organisms.

Escherichia coli 10418

Staphylococcus aureus NCTC 6571

Streptococcus pneumoniae NCTC 12977

Appendix 2 (a) Stata®11.0:Microarray serotyping in children

```
cd "/Users/arox/Desktop/Naor_Arox_Data_Analysis/Arox_Clement_20121217"
log using Arox_Naor_Children_VT_NVT.smcl, replace
use "Arox_Naor_Children_VT_NVT.dta"
```

*** Checking for missing data and categorical variables distribution**

```
Foreach x in vars hiv_status age gender bacterialload cd4p malaria {
tab `x' m
}
bysort hiv_status: sum age, det
di as red "calculates the median age of HIV negative and HIV positive subjects"
ranksum age, by(hiv_status)
di as red "calculates the p-value of the median age of HIV negative and HIV positive
subjects"
logistic vt hiv_status
graph box age,by(vt, title(Age by carriage of 13VT)note /*
*/("Two-sample Wilcoxon rank-sum (Mann-Whitney) test p=0.08""median age
difference 1.6 years"))
graph save Children_box_age_vt.gph, replace
kdensity age if vt==1, addplot (kdensity age if vt==0) title(Age by carriage of/*
*/13VT)note("Two-sample Wilcoxon rank-sum (Mann-Whitney) test
p=0.08""median age difference 1.6 years")
graph save Children_kdensity_age_vt.gph, replace
egen agecat=cut(age),at(0(1)15) label
tab agecat vt
```

```
kdensity agecat if vt==1, addplot (kdensity agecat if vt==0) title(Age by carriage of
13VT)legend(lab(1 13VT carrier) lab(2 non-13VT carrier)) note("Two-sample
Wilcoxon rank-sum (Mann-Whitney) test p=0.08""median age difference 1.6 years")
graph save Children_kdensity_agecat_vt.gph, replace
twoway ///
        (histogram agecat if vt==1, freq disc fc(navy) ylab(,angle(horizontal))) ///
        (histogram agecat if vt==0, freq disc fc(cranberry) ylab(,angle(horizontal)))
///
        , ///
        legend(lab(1 13VT carrier)lab(2 non-13VT carrier)) ///
        title(Age by carriage of 13VT) ///
note("Two-sample Wilcoxon rank-sum (Mann-Whitney) test p=0.08""median age
difference 1.6 years")
        graph save Children_Hist_agecat_vt_1.gph, replace
for each var of varlist a19f - nt1{
        quietly count if missing(`var')
        di as red r(N) as yellow " records missing data in " as green "`var'"
}
```

*** Carriage rate of at least 1 serotype:**

```
gen nocarriage=(a19f==0 & a14==0 & a23f==0 & a6b==0 & a6a==0 & a3==0 &
a19a==0 & ///
                        a4==0 & a18c==0 & a9v==0 & a1==0 & a5==0 & a7f==0 &
a15b==0 &///
                        a13==0 & a16f==0 & a35b==0 & a21==0 & a10a==0 &
a20==0 & a45==0 &///
```

a22f==0 & a15c==0 & a10b==0 & a11d==0 & a9l==0 &

a15a==0 & a35a==0 & ///

a35f==0 & a37==0 & a34==0 & a18a==0 & a16a==0 &

a18f==0 &///

a33a==0 & a7c==0 & a7b==0 & a22a==0 & a33d==0 &

a38==0 & a11a==0 &///

a24b==0 & a2==0 & nt2==0 & nt3b==0 & nt4a==0 &

nt4b==0 & nt1==0)

tab nocarriage

di as red "Dataset includes only carriers"

*** Carriage of 13vPCV Vaccine Types:**

gen VT13=(a19f==1| a14==1| a23f==1| a6a==1| a6b==1| a3==1| a19a==1| a4==1|

a18c==1| a9v==1| a1==1| a5==1| a7f==1)

gen numstgt1=(inrange(numst,2,.))

tab numst numstgt1

lab def multi 0 "single ST" 1 "multiple ST"

lab val numstgt1 multi

ttest age, by(numstgt1)

bysort numstgt1: sum age, det

di as red "calculates the median age of single vs multiple serotype carriers"

ranksum age,by(numstgt1)

di as red "calculates the p-value of the median age of 13VT carriers and non-carriers"

graph box age,by(numstgt1, title(Age by single or multiple carriage)note(t-test/*

*/p=0.33))

di as red "Although mean age for those with multiple carriage is younger by /*

*/0.86 years (10.3 months),"_newline"the differnce is not statistically significant."

graph save Children_age_numstgt1_box.gph, replace

tabstat cd4p, s(mean) by(numst)

ttest cd4p, by(numstgt1)

graph box cd4p,by(numstgt1, title(CD4 percent by single or multiple carriage)note(t-test p=0.47))

graph save Children_cd4p_numstgt1_box.gph, replace

tab numstgt1 gender, col chi

graph pie, over(numstgt1) by(gender, note({&chi}{superscript:2}--test p=0.253)

legend(off) /*

*/title(Single or multiple carriage)subtitle(by gender)) plabel(_all name,

color(white)size(medlarge))

graph save Children_numstgt1_gender_pie.gph, replace

tab numst gender, col chi

graph pie, over(numst) by(gender, note({&chi}{superscript:2}--test p=0.34)

legend(off)title(Number of serotypes carried)subtitle(by gender)) plabel(_all name,

color(white)size(medlarge))

graph save Children_numst_gender_pie.gph, replace

di as red "There is no association between CD4 percent and multiple carriage"

*** Bacterial load by risk factor**

tab bacterialload hiv_status, col

tab bacterialload gender, col

tab bacterialload malaria, col

tabstat age,s(mean) by(bacterialload)

tabstat cd4p,s(mean) by(bacterialload)

```
nptrend cd4p, by(bacterialload)
logistic hiv_status bacterialload
capture logistic age bacterialload
logistic gender bacterialload
di as red "No association between risk factors and bacterial load"
foreach var of varlist a19f - a16a a33a - nt1{
        tab `var' gender, col chi exact
        logistic `var' gender
}
```

*** Serotype specific carriage by HIV status**

```
foreach var of varlist a19f - a16a a33a - nt1{
        tab `var' hiv_status, col chi exact
        logistic `var' hiv_status
}
di as red "ST3 is 11.2 (95%CI 1.3 to 96.6) times more likely in HIV infected,
p=0.03"_newline "Interpret with caution, would probably fail after Bonferroni
correction" _newline"No other ST carriage associated with HIV"
```

*** Distribution graph**

```
preserve
keep sample a19f- nt1
collapse (sum) a19f- nt1
outsheet using st_distribution.csv, c replace
restore
```

*** Cumulative distribution graph**

keep sample a19f- nt1

foreach var of varlist a19f - nt1{

gen `var'dummy=`var'

}

***13VT**

bysort sample: replace a7fdummy=0if

a19fdummy==1|a14dummy==1|a23fdummy==1|a6bdummy==1|a6adummy==1 /*

*/|a3dummy==1|a19adummy==1|a4dummy==1|a18cdummy==1|a9vdummy==1|a1d

ummy==1|a5dummy==1

***invasive potential ST**

bysort sample: replace a2dummy=0if

a19fdummy==1|a14dummy==1|a23fdummy==1|a6bdummy==1|a6adummy==1|a3du

mmy==1 /*

*/|a19adummy==1|a4dummy==1|a18cdummy==1|a9vdummy==1|a1dummy==1|a5d

ummy==1|a7fdummy==1|a15bdummy==1|a13dummy==1

* no invasive potential ST

bysort sample: replace

a38dummy=0ifa19fdummy==1|a14dummy==1|a23fdummy==1|a6bdummy==1|a6ad

ummy==1|a3dummy==1|a19adummy==1|a4dummy==1|a18cdummy==1|a9vdummy

==1|a1dummy==1|a5dummy==1|a7fdummy==1|a15bdummy==1|a13dummy==1

*** nontypeables**

bysort sample: replace nt1dummy=0if

a19fdummy==1|a14dummy==1|a23fdummy==1|a6bdummy==1|a6adummy==1|a3du

mmy==1|a19adummy==1|a4dummy==1|a18cdummy==1|a9vdummy==1|a1dummy=

```
=1|a5dummy==1|a7fdummy==1|a15bdummy==1|a13dummy==1
foreach var of varlist a19fdummy-a15bdummy{
        egen `var'_sum=sum(`var')
        gen `var'_prop=`var'_sum/_N
}
collapse a19fdummy_sum-a15bdummy_prop
foreach var of varlist *_sum { //check here is _sum should be _prop
        gen `var'_cum=`var'
}
```

log close

Appendix 2 (b) Stata®11.0: Microarray serotyping in adults

```
cd "/Users/arox/Desktop/Naor_Arox_Data_Analysis"
log using Arox_Adults_20130416.smcl, replace
use "Arox_Adults_20130416.dta"
* Data edit
drop a9v
replace a28f=0 if missing(a28f)
drop in 118
* Checking for missing data and categorical variables distribution
Foreach x in vars cd4 hiv_status art_usage multi_carriage pcv7_usage {
tab `x', m
* Carriage rate of at least 1 serotype
gen nocarriage=(a3==0 & a19f==0 & a14==0 & a6a==0 & a23f==0 & a6b==0 &
a19a==0 & ///
                a4==0 & a18c==0 & a1==0 & a1==0 & a5==0 & a7f==0 & ///
                a16f==0 & a13==0 & a10b==0 & a21==0 & a35b==0 &
a20==0 & a15b==0 & ///
                a22f==0 & a45==0 & a7b==0 & a11d==0 & a9l==0 &
a35a==0 & a23a==0 & ///
                a28f==0 & a34==0 & a18a==0 & a33a==0 & a38==0 &
a29==0 & a18f==0 &///
                a37==0 & a7c==0 & a46==0 & a31==0 & a12f==0 & a48==0
& nt4a==0 & ///
                nt2==0 & nt4b==0 & nt3b==0)
```

```
tab nocarriage
ttest hiv_status,by(VT7)
```

*** Mean number of serotypes carried**

```
egen numst=rowtotal(a3 - nt3b)
tab numst
sum numst
ci numst
foreach var of varlist hiv_status pcv7_usage art_usage cd4_count{
        di as yellow "T-test of mean number of serotypes carried by `var':"
        ttest numst,by(`var')
        di ""_newline""
}
tab hiv_status pcv7_usage
tab hiv_status numst, col chi
graph bar numst, over(hiv_status) title(Mean number of serotypes carried) /*
*/subtitle(by HIV status) ytitle(mean number of serotypes carried)note(t-test p=0.36)
graph save Adults_hivstat_bar_serotypes.gph, replace
```

*** Serotype specific carriage by HIV status**

```
foreach var of varlist a3-nt3b{
        tab `var' hiv_status, col chi exact
        logistic `var' hiv_status
}
di as red "serotype 13 was frequently isolated from HIV negative than HIV positive
subjects p=0.01"
```

*** Serotype specific carriage by ART usage**

```
foreach var of varlist a3-nt3b{
        tab `var' art_usage, col chi exact
        logistic `var' art_usage
}
```

*** Serotype specific carriage by PCV7 usage**

```
foreach var of varlist a3-nt3b{
        tab `var' pcv7_usage, col chi exact
        logistic `var' pcv7_usage
}
```

*** Serotype specific carriage by CD4 count**

```
foreach var of varlist a3-nt3b{
        tab `var' cd4_count, col chi exact, if hiv_status==1
        logistic `var' cd4_count
}
```

*** Cumulative distribution graph**

```
keep sample a3- nt3b
foreach var of varlist a3 - nt3b{
        gen `var'dummy=`var'
}
```

***13VT**

```
bysort sample: replace a20dummy=0if
a23fdummy==1|a6adummy==1|a19fdummy==1|a14dummy==1|a3dummy==1 /*
*/|a7fdummy==1|a1dummy==1|a6bdummy==1|a4dummy==1|a18cdummy==1|a19a
dummy==1|a5dummy==1
```

***High invasive potential ST**

bysort sample: replace a7bdummy=0if

a23fdummy==1|a6adummy==1|a19fdummy==1|a14dummy==1|a3dummy==1

|a7fdummy==1|a1dummy==1|a6bdummy==1|a4dummy==1|a18cdummy==1|a19adu

mmy==1|a5dummy==1|a20dummy==1|a16fdummy==1|a15bdummy==1|a13dummy

==1|a22fdummy==1|a10bdummy==1|a35bdummy==1|a21dummy==1|a45dummy==

1

*** Low invasive potential ST**

bysort sample: replace nt2dummy=0if

a23fdummy==1|a6adummy==1|a19fdummy==1|a14dummy==1|a3dummy==1|a7fdu

mmy==1|a1dummy==1|a6bdummy==1|a4dummy==1|a18cdummy==1|a19adummy=

=1|a5dummy==1|a20dummy==1|a16fdummy==1|a15bdummy==1|a13dummy==1|a2

2fdummy==1|a10bdummy==1|a35bdummy==1|a21dummy==1|a45dummy==1

*** nontypeables**

bysort sample: replace nt4bdummy=0if

a23fdummy==1|a6adummy==1|a19fdummy==1|a14dummy==1|a3dummy==1|a7fdu

mmy==1|a1dummy==1|a6bdummy==1|a4dummy==1|a18cdummy==1|a19adummy=

=1|a5dummy==1|a20dummy==1|a16fdummy==1|a15bdummy==1|a13dummy==1|a2

2fdummy==1|a10bdummy==1|a35bdummy==1|a21dummy==1|a45dummy==1

di _N

collapse (sum) *dummy

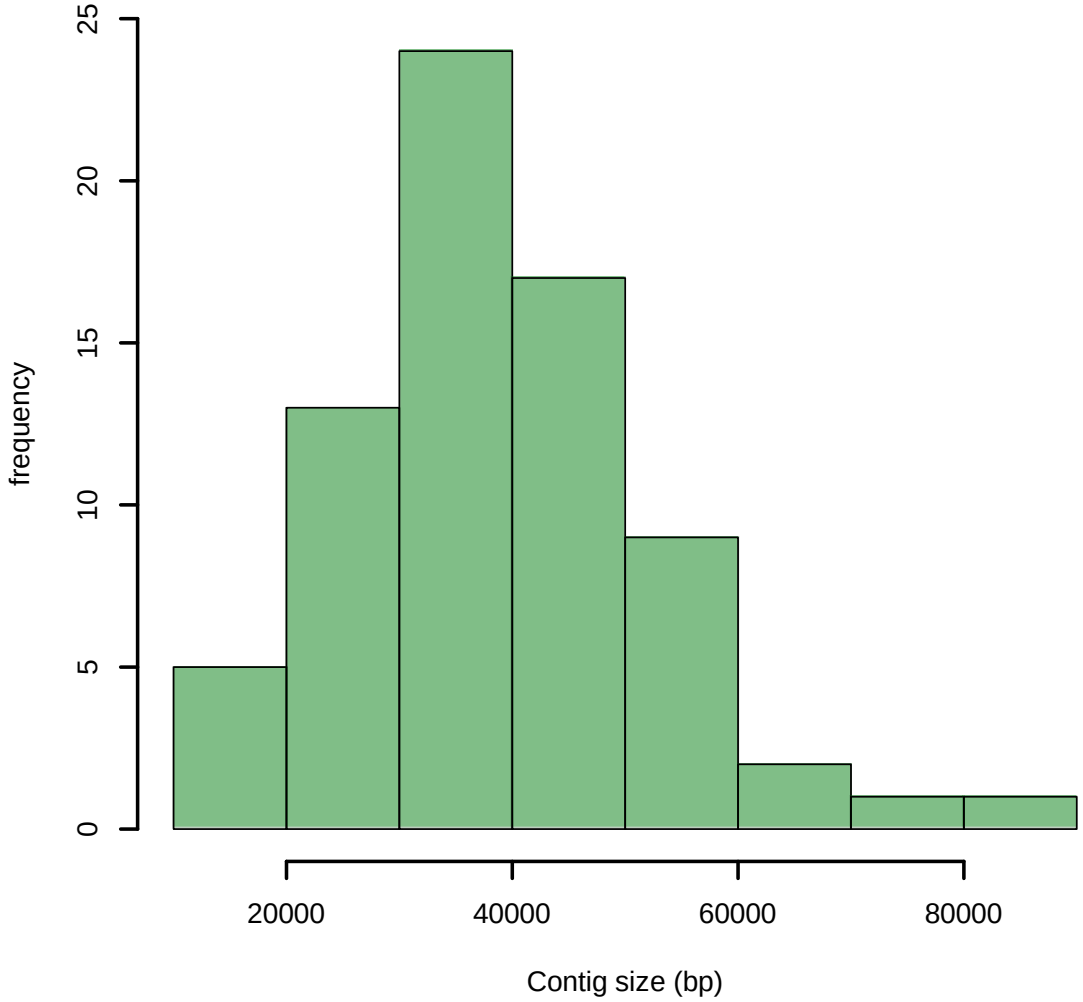
clear

log close

viewArox_Adults_20130416.smcl

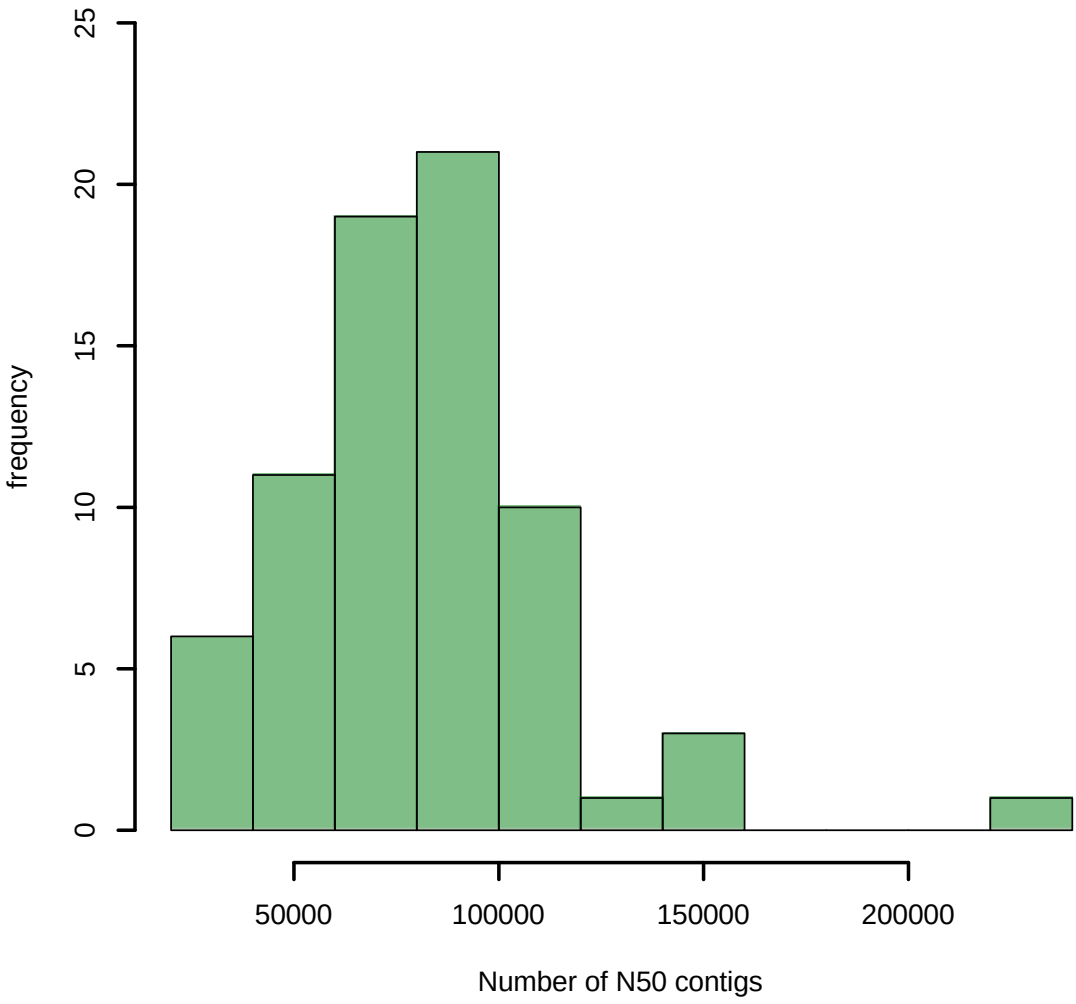
exit

Appendix 3 (a) **Whole genome data: Contig size**



A plot showing the size of contig sequences obtained versus frequency. The size for most of the contigs obtained after assembly ranged from ~20000 to ~60000 base pairs.

Appendix 3 (b) **Whole genome data: Number of N50 contigs**



A plot showing the number of N50 contig sequences obtained versus frequency. The cumulative length form the largest contig to the smallest contig with atleast 50% of the total length (N50 contig number) ranged from ~50000 to ~100000 base pairs.

Appendix 4 Barcoded reverse primers

(sub-tables A-E)

A

Barcode	Sample	Reverse Primer_926R (primer_barcode_adaptor)
GB1	**BC1**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAACGCACGCTAGCCGTCAATTCMTTTRAGT
GB2	**BC2**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAACTCGTCGATGCCGTCAATTCMTTTRAGT
GB3	**BC3**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAACTGTGCGTACCCGTCAATTCMTTTRAGT
GB4	**BC4**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAAGAGATGTCGACCGTCAATTCMTTTRAGT
GB5	**BC5**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAAGCTGCAGTCGCCGTCAATTCMTTTRAGT
GB6	**BC9**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAATCAGTCTCGTCCGTCAATTCMTTTRAGT
GB7	**BC10**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAATCGTGACTCGCCGTCAATTCMTTTRAGT
GB8	**BC11**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACACACTATGGCCCGTCAATTCMTTTRAGT
GB9	**BC12**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACACATGTCTACCCGTCAATTCMTTTRAGT
GB10	**BC18**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACACGAGCCACACCGTCAATTCMTTTRAGT
GB11	**BC19**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACACGGTGTCTACCGTCAATTCMTTTRAGT
GB12	**BC20**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACACTAGATCCGCCGTCAATTCMTTTRAGT
GB13	**BC21**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACACTGTTCATGCCGTCAATTCMTTTRAGT
GB14	**BC22**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACAGACCACTCACCGTCAATTCMTTTRAGT
GB15	**BC23**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACAGAGTCGGCTCCGTCAATTCMTTTRAGT
GB16	**BC27**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACAGCAGTGGTCCCGTCAATTCMTTTRAGT
GB17	**BC29**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACAGCTAGCTTGCCGTCAATTCMTTTRAGT
GB18	**BC30**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACAGTGCTTCATCCGTCAATTCMTTTRAGT
GB19	**BC34**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACAGTTGCGCGACCGTCAATTCMTTTRAGT
GB20	**BC35**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACATCACTTAGCCCGTCAATTCMTTTRAGT
GB21	**P1**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACATGATCGTTCCCGTCAATTCMTTTRAGT
GB22	**P4**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACATGTCACGTGCCGTCAATTCMTTTRAGT
GB23	**P5**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACATTCAGCGCACCGTCAATTCMTTTRAGT
GB24	**P8**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACCACATACATCCCGTCAATTCMTTTRAGT
GB25	**P9**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACCAGACGATGCCCGTCAATTCMTTTRAGT
GB26	**P12**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACCAGCGACTAGCCGTCAATTCMTTTRAGT
GB27	**P13**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACCGCAGAGTCACCGTCAATTCMTTTRAGT

B

Barcode	Sample	Reverse Primer_926R (primer_barcode_adaptor)
GB28	**P16**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACCTCGATCAGACCGTCAATTCMTTTRAGT
GB29	**P17**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACCTGTCTCTCTCCGTCAATTCMTTTRAGT
GB30	**P20**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACGACGTCTTAGCCGTCAATTCMTTTRAGT
GB31	**P21**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACGAGTGCTATCCCGTCAATTCMTTTRAGT
GB32	**P24**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACGATGCGACCACCGTCAATTCMTTTRAGT
GB33	**P25**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACGCAACTGCTACCGTCAATTCMTTTRAGT
GB34	**P28**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACGCGATACTGGCCGTCAATTCMTTTRAGT
GB35	**P29**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACGCGCAGATACCCGTCAATTCMTTTRAGT
GB36	**P32**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACGCTATCTGGACCGTCAATTCMTTTRAGT
GB37	**P33**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACGCTCATGGATCCGTCAATTCMTTTRAGT
GB38	**P36**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACGGATCGTCAGCCGTCAATTCMTTTRAGT
GB39	**P37**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACGGTGAGTGTCCCGTCAATTCMTTTRAGT
GB40	**P40**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACGTACTCAGTGCCGTCAATTCMTTTRAGT
GB41	**P41**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACGTCTGTAGCACCGTCAATTCMTTTRAGT
GB42	**P44**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACGTGAGAGAATCCGTCAATTCMTTTRAGT
GB43	**P45**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACGTGCCGTAGACCGTCAATTCMTTTRAGT
GB44	**P48**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACGTTAGCACACCCGTCAATTCMTTTRAGT
GB45	**P49**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACTACAGCCTATCCGTCAATTCMTTTRAGT
GB46	**P52**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACTACGTGTGGTCCGTCAATTCMTTTRAGT
GB47	**P53**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACTAGCTCCATACCGTCAATTCMTTTRAGT
GB48	**P56**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACTATTGTCACGCCGTCAATTCMTTTRAGT
GB49	**P57**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACTCACGGTATGCCGTCAATTCMTTTRAGT
GB50	**P60**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACTCAGATACTCCCGTCAATTCMTTTRAGT
GB51	**M1**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACTCGATTCGATCCGTCAATTCMTTTRAGT

C

Barcode	Sample	Reverse Primer_926R (primer_barcode_adaptor)
GB52	**M2**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACTCGCACAGGACCGTCAATTCMTTTRAGT
GB53	**M3**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACTCTTCTAGAGCCGTCAATTCMTTTRAGT
GB54	**M4**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACTGACAGCCATCCGTCAATTCMTTTRAGT
GB55	**M5**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACTGATCCTAGTCCGTCAATTCMTTTRAGT
GB56	**M6**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACTGTACGCGTACCGTCAATTCMTTTRAGT
GB57	**M7**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACTGTCGAAGCTCCGTCAATTCMTTTRAGT
GB58	**M10**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACTGTGACTTCACCGTCAATTCMTTTRAGT
GB59	**M11**	CCATCTCATCCCTGCGTGTCTCCGACTCAGACTTGTAGCAGCCCGTCAATTCMTTTRAGT
GB60	**M12**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGAACACGTCTCCCGTCAATTCMTTTRAGT
GB61	**M13**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGACCGTCAGACCCGTCAATTCMTTTRAGT
GB62	**M14**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGACGTGCACTGCCGTCAATTCMTTTRAGT
GB63	**M16**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGACTGCGTACTCCGTCAATTCMTTTRAGT
GB64	**M17**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGAGAGCAAGTGCCGTCAATTCMTTTRAGT
GB65	**M19**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGAGCAAGAGCACCGTCAATTCMTTTRAGT
GB66	**M20**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGAGTAGCTAAGCCGTCAATTCMTTTRAGT
GB67	**M21**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGAGTCCTGAGCCCGTCAATTCMTTTRAGT
GB68	**M22**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGATACACGCGCCCGTCAATTCMTTTRAGT
GB69	**M23**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGATCGGCTCGACCGTCAATTCMTTTRAGT
GB70	**M24**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGATCTCTGCATCCGTCAATTCMTTTRAGT
GB71	**M25**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGATGTTCTGCTCCGTCAATTCMTTTRAGT
GB72	**M27**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCACACCTACACCGTCAATTCMTTTRAGT
GB73	**M29**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCACGAGCCTACCGTCAATTCMTTTRAGT
GB74	**M30**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCAGCACTTGTCCGTCAATTCMTTTRAGT
GB75	**M31**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCAGTCGCGATCCGTCAATTCMTTTRAGT
GB76	**M32**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCATATGAGAGCCGTCAATTCMTTTRAGT
GB77	**M33**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCCATACTGACCCGTCAATTCMTTTRAGT
GB78	**M34**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCGACTGTGCACCGTCAATTCMTTTRAGT
GB79	**M36**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCGAGCTATCTCCGTCAATTCMTTTRAGT
GB80	**M37**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCGCTGATGTGCCGTCAATTCMTTTRAGT

D

Barcode	Sample	Reverse Primer_926R (primer_barcode_adaptor)
GB81	**M38**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCGTAGGTCGTCCGTCAATTCMTTTRAGT
GB82	**M41**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCTATCCACGACCGTCAATTCMTTTRAGT
GB83	**M42**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCTCCATACAGCCGTCAATTCMTTTRAGT
GB84	**M43**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCTCTCAGAGGCCGTCAATTCMTTTRAGT
GB85	**M44**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCTGACTAGTCCCGTCAATTCMTTTRAGT
GB86	**M48**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCTTGACAGCTCCGTCAATTCMTTTRAGT
GB87	**M49**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGGACGCACTGTCCGTCAATTCMTTTRAGT
GB88	**M50**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGGCTACACGACCCGTCAATTCMTTTRAGT
GB89	**M52**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGGTGTGATCGCCCGTCAATTCMTTTRAGT
GB90	**M54**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTACGCTCGAGCCGTCAATTCMTTTRAGT
GB91	**M55**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTACTGCAGGCCCGTCAATTCMTTTRAGT
GB92	**M57**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTAGTATCCTCCCGTCAATTCMTTTRAGT
GB93	**M58**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTCACATCACTCCGTCAATTCMTTTRAGT
GB94	**M59**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTCCATAGCTGCCGTCAATTCMTTTRAGT
GB95	**M60**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTCTACTCTGACCGTCAATTCMTTTRAGT
GB96	**M61**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTCTCGCATATCCGTCAATTCMTTTRAGT
GB97	**M62**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTGAGAGAAGCCCGTCAATTCMTTTRAGT
GB98	**M64**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTGCGATGCGTCCGTCAATTCMTTTRAGT
GB99	**M65**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTGGATGCTCTCCGTCAATTCMTTTRAGT
GB100	**M66**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTGTCACGGTGCCGTCAATTCMTTTRAGT
GB101	**M67**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTGTTCGATCGCCGTCAATTCMTTTRAGT
GB102	**M68**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTTAGTGCGTCCCGTCAATTCMTTTRAGT
GB103	**M69**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTTCAGACGCTCCGTCAATTCMTTTRAGT
GB104	**M70**	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTTCTACGTCACCGTCAATTCMTTTRAGT
GB105	**M71**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATAATCTCGTCGCCGTCAATTCMTTTRAGT
GB106	**K1**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATACACGTGGCGCCGTCAATTCMTTTRAGT
GB107	**K3**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATACAGAGCTCCCCGTCAATTCMTTTRAGT
GB108	**K4**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATACGTCTTCGACCGTCAATTCMTTTRAGT
GB109	**K5**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATACTATTGCGCCCGTCAATTCMTTTRAGT

E

Barcode	Sample	Reverse Primer_926R (primer_barcode_adaptor)
GB110	**K6**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATACTCACTCAGCCGTCAATTCMTTTRAGT
GB111	**K7**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATAGCTCCATACCCGTCAATTCMTTTRAGT
GB112	**K8**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATAGGCGATCTCCCGTCAATTCMTTTRAGT
GB113	**K9**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATATCGCTACTGCCGTCAATTCMTTTRAGT
GB114	**K12**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATATGCCAGTGCCCGTCAATTCMTTTRAGT
GB115	**K16**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATCACGTAGCGGCCGTCAATTCMTTTRAGT
GB116	**K19**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATCACTAGTCACCCGTCAATTCMTTTRAGT
GB117	**K20**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATCAGGCGTGTGCCGTCAATTCMTTTRAGT
GB118	**K21**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATCCGATCACAGCCGTCAATTCMTTTRAGT
GB119	**K24**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATCCTCAGTAGTCCGTCAATTCMTTTRAGT
GB120	**K29**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATCGATCTGTGGCCGTCAATTCMTTTRAGT
GB121	**K35**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATCGCGGACGATCCGTCAATTCMTTTRAGT
GB122	**K36**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATCGCTCGAGGACCGTCAATTCMTTTRAGT
GB123	**K37**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATCGTACAACTCCCGTCAATTCMTTTRAGT
GB124	**K38**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATCTACTACACGCCGTCAATTCMTTTRAGT
GB125	**K40**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATCTCTGGCATACCGTCAATTCMTTTRAGT
GB126	**K43**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATCTGAGCTGGTCCGTCAATTCMTTTRAGT
GB127	**K49**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATCTGGTGCTATCCGTCAATTCMTTTRAGT
GB128	**K54**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATCTTAGACTGCCCGTCAATTCMTTTRAGT
GB129	**K56**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATGACCATCGTGCCGTCAATTCMTTTRAGT
GB130	**K57**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATGACTCATTCGCCGTCAATTCMTTTRAGT
GB131	**K58**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATGAGACTCCACCCGTCAATTCMTTTRAGT
GB132	**K59**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATGATCGAGAGACCGTCAATTCMTTTRAGT
GB133	**K60**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATGCACTGGCGACCGTCAATTCMTTTRAGT
GB134	**K61**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATGCAGCTCAGTCCGTCAATTCMTTTRAGT
GB135	**K62**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATGCCTGAGCAGCCGTCAATTCMTTTRAGT
GB136	**K63**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATGCGTAGTGCGCCGTCAATTCMTTTRAGT
GB137	**K64**	CCATCTCATCCCTGCGTGTCTCCGACTCAGATGGATACGCTCCCGTCAATTCMTTTRAGT

*The V3-V5 region of the 16S rRNA was amplified. The Universal forward primer was 338F:**CCTATCCCCTGTGTGCCTTGGCAGTCTCAGACTCCTACGGGAGGCAG CAG***

Appendix 5 Mothur analysis

(v. 1.30.2)

A. Sequence cleanup/clustering

/software/pathogen/external/apps/usr/bin/mothur

mothur > summary.seqs(fasta=Arox_repeat.fasta)

	Start	End	NBases	Ambigs	Polymer	NumSeqs
Minimum:	1	44	44	0	2	1
2.5%-tile:	1	131	131	0	4	14087
25%-tile:	1	549	549	0	4	140866
Median:	1	552	552	0	4	281732
75%-tile:	1	554	554	0	4	422598
97.5%-tile:	1	567	567	1	5	549377
Maximum:	1	1195	1195	25	31	563463
Mean: 1	538.513		538.513	0.0581014	4.11425	
# of Seqs:	563463					

mothur>trim.seqs(fasta=Arox_repeat.fasta,oligos=Arox_repeat.oligos,qfile=Arox_repeat.qual,maxambig=0,maxhomop=8, flip=T, bdiffs=0, pdiffs=0, qwindowaverage=35, qwindowsize=50, minlength=400)

mothur > summary.seqs(fasta=Arox_repeat.trim.fasta)

	Start	End	NBases	Ambigs	Polymer	NumSeqs
Minimum:	1	400	400	0	4	1
2.5%-tile:	1	401	401	0	4	5369
25%-tile:	1	409	409	0	4	53689
Median:	1	415	415	0	4	107378
75%-tile:	1	425	425	0	4	161066
97.5%-tile:	1	483	483	0	5	209386
Maximum:	1	531	531	0	7	214754
Mean: 1	420.08	420.08	0	4.03635		
# of Seqs:	214754					

mothur > unique.seqs(fasta=Arox_repeat.trim.fasta)

mothur>summary.seqs(fasta=Arox_repeat.trim.unique.fasta,name=Arox_repeat.trim.names)

	Start	End	NBases	Ambigs	Polymer	NumSeqs
Minimum:	1	400	400	0	4	1
2.5%-tile:	1	401	401	0	4	5369
25%-tile:	1	409	409	0	4	53689
Median:	1	415	415	0	4	107378
75%-tile:	1	425	425	0	4	161066
97.5%-tile:	1	483	483	0	5	209386
Maximum:	1	531	531	0	7	214754
Mean: 1	420.08	420.08	0	4.03635		

mothur>align.seqs(fasta=Arox_repeat.trim.unique.fasta,reference=silva.bacteria.fasta)

mothur > summary.seqs(fasta=Arox_repeat.trim.unique.align)

	Start	End	NBases	Ambigs	Polymer	NumSeqs
Minimum:	7705	27656	400	0	4	1
2.5%-tile:	8507	27659	401	0	4	672
25%-tile:	10239	27659	409	0	4	6714
Median:	10275	27659	417	0	4	13427
75%-tile:	10351	27659	429	0	4	20140
97.5%-tile:	10369	27659	490	0	6	26182
Maximum:	11890	28426	531	0	7	26853
Mean: 10190.6		27658.7	422.86	0	4.11958	
# of Seqs:	26853					

mothur>screen.seqs(fasta=Arox_repeat.trim.unique.align,start=10370,end=27656, maxhomop=6,name=Arox_repeat.trim.names,group=Arox_repeat.groups,qfile=Arox_repeat.qual, alignreport=Arox_repeat.trim.unique.align.report)

mothur>summary.seqs(fasta=Arox_repeat.trim.unique.good.align,name=Arox_repeat.trim.good.names)

	Start	End	NBases	Ambigs	Polymer	NumSeqs
Minimum:	7705	27656	400	0	4	1
2.5%-tile:	9804	27659	401	0	4	5366
25%-tile:	10247	27659	409	0	4	53654
Median:	10279	27659	415	0	4	107308
75%-tile:	10351	27659	425	0	4	160962
97.5%-tile:	10368	27659	483	0	5	209250
Maximum:	10370	28426	531	0	6	214615
Mean:	10229.7	27649.7	420.092	0	4.03632	
# of unique seqs:	26815; total # of seqs: 214615					

mothur>unique.seqs(fasta=Arox_repeat.trim.unique.good.filter.fasta,name=Arox_repeat.trim.good.names)

mothur>summary.seqs(fasta=Arox_repeat.trim.unique.good.filter.unique.fasta,name=Arox_repeat.trim.unique.good.filter.names)

	Start	End	NBases	Ambigs	Polymer	NumSeqs
Minimum:	1	772	392	0	4	1
2.5%-tile:	1	772	398	0	4	5366
25%-tile:	1	772	398	0	4	53654
Median:	1	772	398	0	4	107308
75%-tile:	1	772	398	0	4	160962
97.5%-tile:	1	772	399	0	5	209250
Maximum:	1	772	404	0	6	214615
Mean:	1	769.62	399.189	0	4.03629	
# of unique seqs:	8135					
total # of seqs:	214615					

mothur>pre.cluster(fasta=Arox_repeat.trim.unique.good.filter.unique.fasta,name=Arox_repeat.trim.unique.good.filter.names, diffs=1)

Mothur>summary.seqs(fasta=Arox_repeat.trim.unique.good.filter.unique.precluster.fasta, name=Arox_repeat.trim.unique.good.filter.unique.precluster.names)

	Start	End	NBases	Ambigs	Polymer	NumSeqs
Minimum:	1	772	392	0	4	1
2.5%-tile:	1	772	398	0	4	5366
25%-tile:	1	772	398	0	4	53654
Median:	1	772	398	0	4	107308
75%-tile:	1	772	398	0	4	160962
97.5%-tile:	1	772	399	0	5	209250
Maximum:	1	772	404	0	6	214615
Mean: 1	769.62	399.192	0	4.03337		

# of unique seqs: 3918
total # of seqs: 214615

mothur>chimera.perseus(fasta=Arox_repeat.trim.unique.good.filter.unique.precluster.fasta, name=Arox_repeat.trim.unique.good.filter.unique.precluster.names)

mothur>remove.seqs(accnos=Arox_repeat.trim.unique.good.filter.unique.precluster.perseus.accnos,fasta=Arox_repeat.trim.unique.good.filter.unique.precluster.fasta,name=Arox_repeat.trim.unique.good.filter.unique.precluster.names,group=Arox_repeat.good.groups)

mothur>summary.seqs(fasta=Arox_repeat.trim.unique.good.filter.unique.precluster.pick.fasta, name=Arox_repeat.trim.unique.good.filter.unique.precluster.pick.names)

	Start	End	NBases	Ambigs	Polymer	NumSeqs
Minimum:	1	772	392	0	4	1
2.5%-tile:	1	772	398	0	4	5357
25%-tile:	1	772	398	0	4	53565
Median:	1	772	398	0	4	107130
75%-tile:	1	772	398	0	4	160694
97.5%-tile:	1	772	399	0	5	208902
Maximum:	1	772	404	0	6	214258
Mean: 1	769.623	399.191	0	4.03307		

# of unique seqs: 3747
total # of seqs: 214258

mothur>dist.seqs(fasta=Arox_repeat.trim.unique.good.filter.unique.precluster.pick.fasta)

mothur>cluster(column=Arox_repeat.trim.unique.good.filter.unique.precluster.pick.dist, name=Arox_repeat.trim.unique.good.filter.unique.precluster.pick.names)

mothur>make.shared(list=Arox_repeat.trim.unique.good.filter.unique.precluster.pick.an.list, group=Arox_repeat.good.pick.groups, label=0.03)

B. Alpha Diversity (Output files plotted in Excel/Stata v.11)

Using file name: Arox_repeat_clean.shared

mothur > sub.sample(shared=Arox_repeat_clean.shared, size=1000)

Output File Names: Arox_repeat_clean.0.03.subsample.shared

mothur>collect.single(shared=Arox_repeat_clean.shared,calc=chao-invsimpson, freq=5)

mothur>collect.single(shared=Arox_repeat_clean.0.03.subsample.shared, calc=chao-invsimpson, freq=5)

mothur>rarefaction.single(shared=Arox_repeat_clean.shared, calc=sobs, freq=100)

mothur>summary.single(shared=Arox_repeat_clean.0.03.subsample.shared,calc=nseqs-coverage-sobs-invsimpson)

C. Determining species diversity by Age, HIV status, PCV and ART Usage using Venn command in mothur

Generating Venn diagrams in mothur to compare species distribution.For example, determining species diversity between HIV positive and HIV negative children.*Data processing was as follows:*

1. Making a copy of the 'subsampled.shared' file.
2. Removing all adults' samples.
3. Removing all the duplicated samples from longitudinal data set, so that there is only one sample per child.
4. All HIV positive and negative columns were added together and called 'HIVpos' and 'HIVneg' respectively.
5. This generated a shared file (Kids_HIV.shared) with only two samples in it. A tab-delimited format of the shared file was used to generate Venn diagrams in mothur.
6. Using mothur to generate Venn diagram

mothur > venn(shared=Kids_HIV.shared)

The scaleable vector graphic (svg) file produced can be viewed in Firefox, Safari or any other browser. The above procedure was repeated to determine species diversity for other parameters such as Age, PCV and ART usage.

D. Principal component analysis using Bray Curtis and Jaccard

Bray Curtis is employed to determine the type of species present in a given sample and their proportions, while Jaccard determines only the species present (Porat et al.).

mothur>dist.shared(shared=Arox_repeat_clean.0.03.subsample.shared,calc=braycurtis-jclass)

mothur > pcoa(phylip=Arox_repeat_clean.0.03.subsample.braycurtis.0.03.lt.dist)

mothur > pcoa(phylip=Arox_repeat_clean.0.03.subsample.jclass.0.03.lt.dist)

mothur >metastats(shared=Adults_children.shared, design=Adults_children.design)

mothur > metastats(shared=Child_BT_KA.shared, design=Child_BT_KA.design)

Appendix 6 Multiple carriage in children

Sample	Microarray Serotypes (Relative abundance)				Number of serotypes
	Sero1 (%)	Sero2 (%)	Sero3 (%)	Sero4 (%)	
A-718*	19F (95)	NT4a (5)			2
B-439	16F (99)	NT4a (1)			2
C-710	NT3b (41)	35B (37)	NT2 (22)		3
D-927*	19F (93)	35F (7)			2
E-117	15B (98)	16F (2)			2
F-607	10B (51)	NT4b (44)	34 (5)		3
G-940*	21 (68)	6B (32)			2
H-539*	23F (54)	15B (33)	14 (12)	NT3b (1)	4
I-067	15B (96)	11D (4)			2
J-808*	23F (54)	18A (39)	6B (7)		3
K-086*	3 (57)	34 (29)	NT4a (14)		3
L-556*	15B (72)	6A (28)			2
M-651*	3 (99)	35A (1)			2
N-516*	10B (89)	19A (6)	45 (5)		3
O-535*	13 (89)	3 (11)			2
P-590	NT3b (86)	10B (14)			2
Q-650*	23F (98)	19F (2)			2
R-636*	NT3b (73)	4 (17)	35B (11)		3
S-698	2 (55)	NT4b (45)			2
T-697	10B (87)	NT3b (13)			2
U-656	35B (83)	13 (17)			2
V-587*	19F (99)	16A (1)			2
X-1Y7*	1 (42)	14 (37)	3 (21)		3
Y-2D4	15B (99)	18F (1)			2
Z-2IH*	1 (95)	NT2 (5)			2
AA-064	15B (99)	9L (1)			2
AB-1YA*	15C (48)	14 (46)	38 (6)		3
AC-2C2*	NT3b (66)	1 (19)	NT2 (16)		3
AD-2DR*	NT3b (53)	18C (47)			2
AE-22U*	6B (63)	NT2 (37)			2
AF-238*	13 (44)	NT2 (36)	14 (20)		3
AG-728*	3 (57)	15B (43)			2
AH-1VK*	15B (50)	5 (44)	6B (6)		3
AI-1YH*	23F (87)	20 (13)			2
AJ-1YZ*	13 (77)	6B (23)			2
AK-22W*	7C (89)	NT4a (5)	37 (5)	14 (1)	4
AL-122*	9L (70)	19F (27)	14 (3)		3
AM-2M2*	23F (51)	9V (26)	11A (23)		3
AN-2M9*	14 (96)	7F (4)			2
AO-811*	19F (71)	23F (29)			2
AP-693*	6A (92)	9V (8)			2
AQ-2M3*	6A (91)	NT2 (9)			2
AR-1XY*	15A (65)	6A (35)			2
AS-1VS*	19A (59)	4 (41)			2
AT-235*	19F (67)	33D (18)			2
AU-568*	19F (67)	6B (33)			2

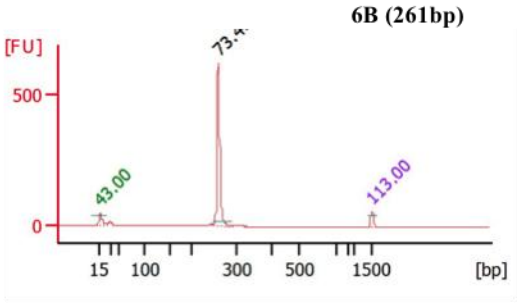
Samples involving a vaccine type in multiple carriage are marked with a (*).

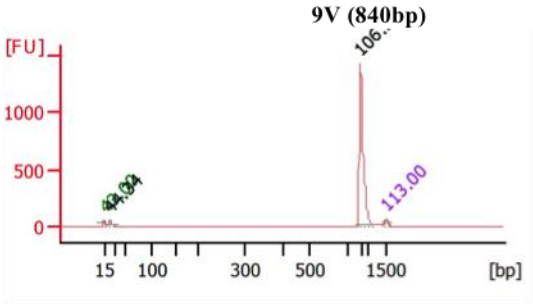
Appendix 7 (a) Multiplex PCR serotyping primers

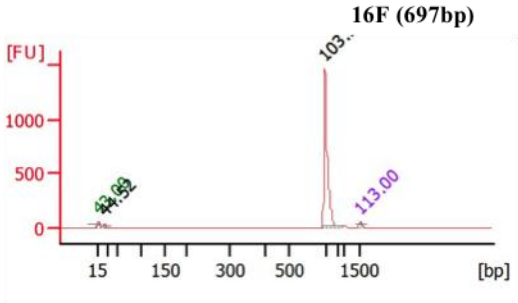
Serotype	Primers	Nucleotide Position	Product size (bp)	Gene target	Accesion no.	Reference
6B	AAT TTG TAT TTT ATT CAT GCC TAT ATC TGG -f	8656	250	wciP	CR931639	(Pai et al., 2006)
6B	TTA GCG GAG ATA ATT TAA AAT GAT GAC TA -r	8877				(Pai et al., 2006)
9L	GAA CTG AAT AAG TCA GAT TTA ATC AGC -f	11948	516	wzx	CR931647	(Dias et al., 2007)
9L	ACC AAG ATC TGA CGG GCT AAT CAA T -r	12439				(Dias et al., 2007)
9V	GGG TTC AAA G TC AGA CAG TG A ATC TTA A -f	9966	816	wzy	CR931648	(da Gloria Carvalho et al., 2010)
9V	CCA TGA ATG A AA TCA ACA TT G TCA GTA GC -r	10753				(da Gloria Carvalho et al., 2010)
10A	GGT GTA GAT TTA CCA TTA GTG TCG GCA GAC -f	12423	628	wcrG	CR931649	(Pai et al., 2006)
10A	GAA TTT CTT CTT TAA GAT TCG GAT ATT TCT C -r	13021				(Pai et al., 2006)
11D	GGA CAT GTT CAG GTG ATT TCC CAA TAT AGT G -f	11640	463	wzy	CR931653	(Pai et al., 2006)
11D	GAT TAT GAG TGT AAT TTA TTC CAA CTT CTC CC -r	12071				(Pai et al., 2006)
14	GAA ATG TTA CTT GGC GCA GGT GTC AGA ATT -f	7959	189	wzy	CR931662	(Dias et al., 2007)
14	GCC AAT ACT TCT TAG TCT CTC AGA TGA AT -r	8119				(Dias et al., 2007)
16F	GAA TTT TTC AGG CGT GGG TGT TAA AAG -f	11679	716	wzy	CR931668	(da Gloria Carvalho et al., 2010)
16F	CAG CAT ATA GCA CCG CTA AGC AAA TA -r	12371				(da Gloria Carvalho et al., 2010)
19F	GTT AAG ATT GCT GAT CGA TTA ATT GAT ATC C -f	11135	304	wzy	CR931678	(Pai et al., 2006)
19F	GTA ATA TGT CTT TAG GGC GTT TAT GGC GAT AG -r	11407				(Pai et al., 2006)
23B	CCA CAA TTA G CG CTA TAT TCA TTC AAT CG -f	13227	199	wzx	CR931684	(da Gloria Carvalho et al., 2010)
23B	GTC CAC GCT GAA TAA AAT GAA GCT CCG -r	13399				(da Gloria Carvalho et al., 2010)
23F	GTA ACA GTT GCT GTA GAG GGA ATT GGC TTT TC -f	8768	384	wzy	CR931685	(Pai et al., 2006)
23F	CAC AAC ACC TAA CAC TCG ATG GCT ATA TGA TTC -f	9119				(Pai et al., 2006)
35F	GAA CAT AGT CGC TAT TGT ATT TTA TTT AAA GCA A -f	7374	517	wzy	CR931707	(Pai et al., 2006)
35F	GAC TAG GAG CAT TAT TCC TAG AGC GAG TAA ACC -r	7858				(Pai et al., 2006)

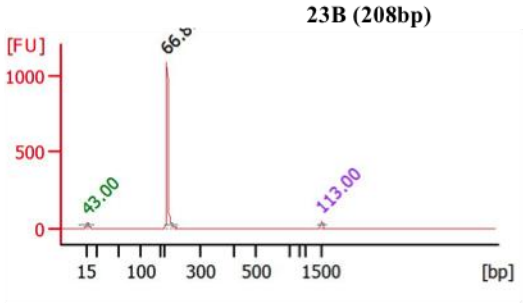
Appendix 7 (b) Bioanalyser electropherograms

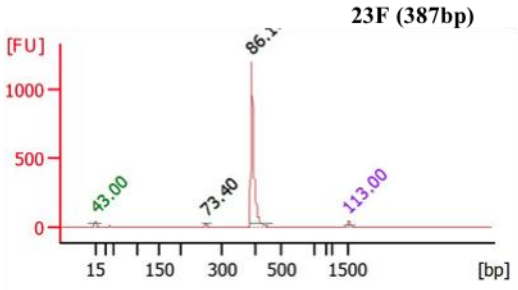
Confirmed the presence 6B (250bp, wciP), 9V (816bp, wzy), 16F (716bp, wzy), 23B (199bp, wzx) and 23F (384bp, wzy). This confirmed the presence of 23B (199bp, wzx) and 35F (517bp, wzy).

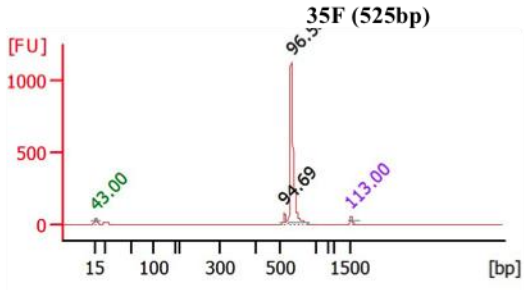












Appendix 8 (a) OTU distribution in children and adults

OTU	mean(children)	mean(adults)	p-value
Otu085	0.000000	0.000056	0.000999
Otu107	0.000000	0.000444	0.000999
Otu002	0.000000	0.000111	0.000999
Otu029	0.000000	0.000222	0.000999
Otu144	0.000000	0.000056	0.000999
Otu126	0.000000	0.000056	0.000999
Otu167	**0.062950**	**0.010111**	**0.027972**
Otu137	0.002600	0.000056	0.152847
Otu117	0.000000	0.000111	0.187812
Otu076	0.000400	0.000000	0.209790
Otu149	0.001950	0.005444	0.212787
Otu154	0.000050	0.000944	0.264735
Otu160	0.010750	0.003222	0.283716
Otu161	0.000950	0.000222	0.287712
Otu170	0.000350	0.001389	0.391608
Otu142	0.000125	0.001944	0.396603
Otu005	0.000075	0.000000	0.412587
Otu159	0.000625	0.003556	0.446553
Otu080	0.000600	0.000056	0.462537
Otu109	0.000500	0.000056	0.483516
Otu061	0.000050	0.000000	0.498501
Otu014	0.000050	0.000000	0.500500
Otu011	0.000100	0.000000	0.502498
Otu164	0.000050	0.000056	0.628372
Otu155	0.003850	0.032389	0.640360
Otu103	0.000825	0.003111	0.685315
Otu133	0.000025	0.000056	0.713287
Otu006	0.000050	0.000722	0.760240
Otu163	0.000075	0.018222	0.778222
Otu021	0.000025	0.000000	0.782218
Otu106	0.002850	0.001000	0.786214
Otu132	0.001350	0.000167	0.790210
Otu172	0.003525	0.005111	0.791209
Otu165	0.021400	0.038167	0.797203
Otu166	0.000175	0.000222	0.858142
Otu017	0.000050	0.000000	0.881119
Otu115	0.000100	0.000000	0.901099
Otu173	0.881200	0.871833	0.904096
Otu090	0.000075	0.000000	0.912088
Otu171	0.002250	0.000389	0.984016

Appendix 8 (b) OTU distribution in children

Otu	Karonga (mean)	Blantyre (mean)	p-value
Otu172	0.000000	0.005875	**0.000999**
Otu115	0.000000	0.000167	**0.000999**
Otu006	0.000000	0.000083	**0.000999**
Otu017	0.000000	0.000083	**0.000999**
Otu167	0.148688	0.005792	**0.001998**
Otu137	0.006500	0.000000	**0.008991**
Otu160	0.000188	0.017792	**0.009990**
Otu173	0.788375	0.943083	**0.023976**
Otu165	0.047125	0.004250	**0.029970**
Otu163	0.000000	0.000125	**0.048951**
Otu106	0.000000	0.004750	0.052947
Otu142	0.000000	0.000208	0.097902
Otu103	0.001500	0.000375	0.148851
Otu154	0.000000	0.000083	0.207792
Otu159	0.001125	0.000292	0.226773
Otu021	0.000000	0.000042	0.305694
Otu061	0.000125	0.000000	0.312687
Otu011	0.000250	0.000000	0.314685
Otu014	0.000125	0.000000	0.331668
Otu005	0.000188	0.000000	0.342657
Otu161	0.001625	0.000500	0.368631
Otu132	0.000000	0.002250	0.413586
Otu109	0.001125	0.000083	0.496503
Otu164	0.000063	0.000042	0.507493
Otu149	0.001250	0.002417	0.509491
Otu170	0.000250	0.000417	0.548452
Otu133	0.000000	0.000042	0.691309
Otu166	0.000063	0.000250	0.698302
Otu080	0.001313	0.000125	0.720280
Otu155	0.000063	0.006375	0.763237
Otu076	0.000063	0.000625	0.820180
Otu090	0.000000	0.000125	0.823177
Otu171	0.000000	0.003750	0.926074

Appendix 8 (b) OTU distribution in children

Otu	Karonga (mean)	Blantyre (mean)	p-value
Otu172	0.000000	0.005875	**0.000999**
Otu115	0.000000	0.000167	**0.000999**
Otu006	0.000000	0.000083	**0.000999**
Otu017	0.000000	0.000083	**0.000999**
Otu167	0.148688	0.005792	**0.001998**
Otu137	0.006500	0.000000	**0.008991**
Otu160	0.000188	0.017792	**0.009990**
Otu173	0.788375	0.943083	**0.023976**
Otu165	0.047125	0.004250	**0.029970**
Otu163	0.000000	0.000125	**0.048951**
Otu106	0.000000	0.004750	0.052947
Otu142	0.000000	0.000208	0.097902
Otu103	0.001500	0.000375	0.148851
Otu154	0.000000	0.000083	0.207792
Otu159	0.001125	0.000292	0.226773
Otu021	0.000000	0.000042	0.305694
Otu061	0.000125	0.000000	0.312687
Otu011	0.000250	0.000000	0.314685
Otu014	0.000125	0.000000	0.331668
Otu005	0.000188	0.000000	0.342657
Otu161	0.001625	0.000500	0.368631
Otu132	0.000000	0.002250	0.413586
Otu109	0.001125	0.000083	0.496503
Otu164	0.000063	0.000042	0.507493
Otu149	0.001250	0.002417	0.509491
Otu170	0.000250	0.000417	0.548452
Otu133	0.000000	0.000042	0.691309
Otu166	0.000063	0.000250	0.698302
Otu080	0.001313	0.000125	0.720280
Otu155	0.000063	0.006375	0.763237
Otu076	0.000063	0.000625	0.820180
Otu090	0.000000	0.000125	0.823177
Otu171	0.000000	0.003750	0.926074